Graduate Texts in Mathematics

Neal Koblitz

Introduction to Elliptic Curves and Modular Forms

Second Edition

Springer

Graduate Texts in Mathematics 97

Springer-Science+Business Media, LLC



Neal Koblitz

Introduction to Elliptic Curves and Modular Forms

Second Edition

With 24 Illustrations

Springer

Neal Koblitz
Department of Mathematics
University of Washington
Seattle, WA 98195
USA

Mathematics Subject Classification (2000): 11-01, 11Dxx, 11Gxx, 11Rxx, 14H45

Library of Congress Cataloging-in-Publication Data
Koblitz, Neal.
Introduction to elliptic curves and modular forms / Neal Koblitz.
— 2nd ed.
p. cm. — (Graduate texts in mathematics; 97)
ISBN 978-1-4612-6942-7 ISBN 978-1-4612-0909-6 (eBook)
DOI 10.1007/978-1-4612-0909-6
1. Curves, Elliptic. 2. Forms, Modular. 3. Number Theory.
I. Title. II. Series.
QA567.2E44K63 1993
516.3′52—dc20 92-41778

Printed on acid-free paper.

Typeset by Asco Trade Typesetting Ltd., Hong Kong.

9 8 7 6 5 4

Preface to the First Edition

This textbook covers the basic properties of elliptic curves and modular forms, with emphasis on certain connections with number theory. The ancient "congruent number problem" is the central motivating example for most of the book.

My purpose is to make the subject accessible to those who find it hard to read more advanced or more algebraically oriented treatments. At the same time I want to introduce topics which are at the forefront of current research. Down-to-earth examples are given in the text and exercises, with the aim of making the material readable and interesting to mathematicians in fields far removed from the subject of the book.

With numerous exercises (and answers) included, the textbook is also intended for graduate students who have completed the standard first-year courses in real and complex analysis and algebra. Such students would learn applications of techniques from those courses, thereby solidifying their understanding of some basic tools used throughout mathematics. Graduate students wanting to work in number theory or algebraic geometry would get a motivational, example-oriented introduction. In addition, advanced undergraduates could use the book for independent study projects, senior theses, and seminar work.

This book grew out of lecture notes for a course I gave at the University of Washington in 1981–1982, and from a series of lectures at the Hanoi Mathematical Institute in April, 1983. I would like to thank the auditors of both courses for their interest and suggestions. My special gratitude is due to Gary Nelson for his thorough reading of the manuscript and his detailed comments and corrections. I would also like to thank Professors J. Buhler, B. Mazur, B. H. Gross, and Huynh Mui for their interest, advice and encouragement.

The frontispiece was drawn by Professor A. T. Fomenko of Moscow State University to illustrate the theme of this book. It depicts the family of elliptic curves (tori) that arises in the congruent number problem. The elliptic curve corresponding to a natural number n has branch points at $0, \infty, n$ and $-n$. In the drawing we see how the elliptic curves interlock and deform as the branch points $\pm n$ go to infinity.

Note: References are given in the form [Author year]; in case of multiple works by the same author in the same year, we use a, b, ... after the date to indicate the order in which they are listed in the Bibliography.

Seattle, Washington NEAL KOBLITZ

Preface to the Second Edition

The decade since the appearance of the first edition has seen some major progress in the resolution of outstanding theoretical questions concerning elliptic curves. The most dramatic of these developments have been in the direction of proving the Birch and Swinnerton-Dyer conjecture. Thus, one of the changes in the second edition is to update the bibliography and the discussions of the current state of knowledge of elliptic curves.

It was also during the 1980s that, for the first time, several important practical applications were found for elliptic curves. In the first place, the algebraic geometry of elliptic curves (and other algebraic curves, especially the curves that parametrize modular forms) were found to provide a source of new error-correcting codes which sometimes are better in certain respects than all previously known ones (see [van Lint 1988]). In the second place, H.W. Lenstra's unexpected discovery of an improved method of factoring integers based on elliptic curves over finite fields (see [Lenstra 1987]) led to a sudden interest in elliptic curves among researchers in cryptography. Further cryptographic applications arose as the groups of elliptic curves were used as the "site" of so-called "public key" encryption and key exchange schemes (see [Koblitz 1987], [Miller 1986], [Menezes and Vanstone 1990]).

Thus, to a much greater extent than I would have expected when I wrote this book, readers of the first edition came from applied areas of the mathematical sciences as well as the more traditional fields for the study of elliptic curves, such as algebraic geometry and algebraic number theory.

I would like to thank the many readers who suggested corrections and improvements that have been incorporated into the second edition.

Contents

CHAPTER I

From Congruent Numbers to Elliptic Curves

The theory of elliptic curves and modular forms is one subject where the most diverse branches of mathematics come together: complex analysis, algebraic geometry, representation theory, number theory. While our point of view will be number theoretic, we shall find ourselves using the type of techniques that one learns in basic courses in complex variables, real variables, and algebra. A well-known feature of number theory is the abundance of conjectures and theorems whose statements are accessible to high school students but whose proofs either are unknown or, in some cases, are the culmination of decades of research using some of the most powerful tools of twentieth century mathematics.

We shall motivate our choice of topics by one such theorem: an elegant characterization of so-called "congruent numbers" that was proved by J. Tunnell [Tunnell 1983]. A few of the proofs of necessary results go beyond our scope, but many of the ingredients in the proof of Tunnell's theorem will be developed in complete detail.

Tunnell's theorem gives an almost complete answer to an ancient problem: find a simple test to determine whether or not a given integer n is the area of some right triangle all of whose sides are rational numbers. A natural number n is called "congruent" if there exists a right triangle with all three sides rational and area n. For example, 6 is the area of the 3–4–5 right triangle, and so is a congruent number.

Right triangles whose sides are integers X, Y, Z (a "Pythagorean triple") were studied in ancient Greece by Pythagoras, Euclid, Diophantus, and others. Their central discovery was that there is an easy way to generate all such triangles. Namely, take any two positive integers a and b with $a > b$, draw the line in the uv-plane through the point $(-1, 0)$ with slope b/a. Let (u, v) be the second point of intersection of this line with the unit circle (see Fig. I.1). It is not hard to show that

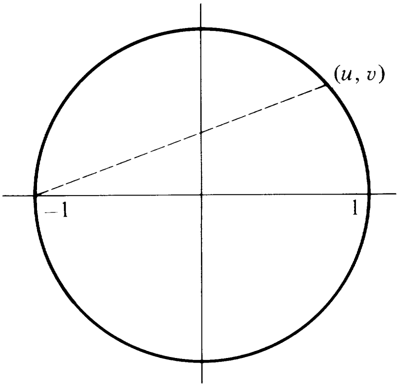


Figure I.1

$$u = \frac{a^2 - b^2}{a^2 + b^2}, \qquad v = \frac{2ab}{a^2 + b^2}.$$

Then the integers $X = a^2 - b^2$, $Y = 2ab$, $Z = a^2 + b^2$ are the sides of a right triangle; the fact that $X^2 + Y^2 = Z^2$ follows because $u^2 + v^2 = 1$. By letting a and b range through all positive integers with $a > b$, one gets all possible Pythagorean triples (see Problem 1 below).

Although the problem of studying numbers n which occur as areas of rational right triangles was of interest to the Greeks in special cases, it seems that the congruent number problem was first discussed systematically by Arab scholars of the tenth century. (For a detailed history of the problem of determining which numbers are "congruent", see [L. E. Dickson 1952, Ch. XVI]; see also [Guy 1981, Section D27].) The Arab investigators preferred to rephrase the problem in the following equivalent form: given n, can one find a rational number x such that $x^2 + n$ and $x^2 - n$ are both squares of rational numbers? (The equivalence of these two forms of the congruent number problem was known to the Greeks and to the Arabs; for a proof of this elementary fact, see Proposition 1 below.)

Since that time, some well-known mathematicians have devoted considerable energy to special cases of the congruent number problem. For example, Euler was the first to show that $n = 7$ is a congruent number. Fermat showed that $n = 1$ is *not*; this result is essentially equivalent to Fermat's Last Theorem for the exponent 4 (i.e., the fact that $X^4 + Y^4 = Z^4$ has no nontrivial integer solutions).

It eventually became known that the numbers 1, 2, 3, 4 are not congruent numbers, but 5, 6, 7 are. However, it looked hopeless to find a straightforward criterion to tell whether or not a given n is congruent. A major advance in the twentieth century was to place this problem in the context of the arithmetic theory of elliptic curves. It was in this context that Tunnell was able to prove his remarkable theorem.

Here is part of what Tunnell's theorem says (the full statement will be given later):

Theorem (Tunnell). *Let n be an odd squarefree natural number. Consider the two conditions:*

(A) *n is congruent;*
(B) *the number of triples of integers (x, y, z) satisfying $2x^2 + y^2 + 8z^2 = n$ is equal to twice the number of triples satisfying $2x^2 + y^2 + 32z^2 = n$.*

Then (A) *implies* (B)*; and, if a weak form of the so-called Birch–Swinnerton-Dyer conjecture is true, then* (B) *also implies* (A).

The central concepts in the proof of Tunnell's theorem—the Hasse–Weil L-function of an elliptic curve, the Birch–Swinnerton-Dyer conjecture, modular forms of half integer weight—will be discussed in later chapters. Our concern in this chapter will be to establish the connection between congruent numbers and a certain family of elliptic curves, in the process giving the definition and some basic properties of elliptic curves.

§1. Congruent numbers

Let us first make a more general definition of a congruent number. A positive rational number $r \in \mathbb{Q}$ is called a "congruent number" if it is the area of some right triangle with rational sides. Suppose r is congruent, and $X, Y, Z \in \mathbb{Q}$ are the sides of a triangle with area r. For any nonzero $r \in \mathbb{Q}$ we can find some $s \in \mathbb{Q}$ such that s^2r is a squarefree integer. But the triangle with sides sX, sY, sZ has area s^2r. Thus, without loss of generality we may assume that $r = n$ is a squarefree natural number. Expressed in group language, we can say that whether or not a number r in the multiplicative group $\mathbb{Q}^+$ of positive rational numbers has the congruent property depends only on its coset modulo the subgroup $(\mathbb{Q}^+)^2$ consisting of the squares of rational numbers; and each coset in $\mathbb{Q}^+/(\mathbb{Q}^+)^2$ contains a unique squarefree natural number $r = n$. In what follows, when speaking of congruent numbers, we shall always assume that the number is a squarefree positive integer.

Notice that the definition of a congruent number does not require the sides of the triangle to be integral, only rational. While $n = 6$ is the smallest possible area of a right triangle with integer sides, one can find right triangles with rational sides having area $n = 5$. The right triangle with sides $1\frac{1}{2}$, $6\frac{2}{3}$, $6\frac{5}{6}$ is such a triangle (see Fig. I.2). It turns out that $n = 5$ is the smallest congruent number (recall that we are using "congruent number" to mean "congruent squarefree natural number").

There is a simple algorithm using Pythagorean triples (see the problems below) that will eventually list all congruent numbers. Unfortunately, given

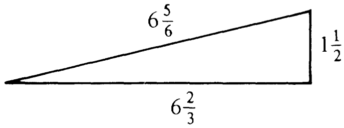


Figure I.2

n, one cannot tell how long one must wait to get n if it is congruent; thus, if n has not appeared we do not know whether this means that n is not a congruent number or that we have simply not waited long enough. From a practical point of view, the beauty of Tunnell's theorem is that his condition (B) can be easily and rapidly verified by an effective algorithm. Thus, his theorem almost settles the congruent number problem, i.e., the problem of finding a verifiable criterion for whether a given n is congruent. We must say "almost settles" because in one direction the criterion is only known to work in all cases if one assumes a conjecture about elliptic curves.

Now suppose that X, Y, Z are the sides of a right triangle with area n. This means: $X^2 + Y^2 = Z^2$, and $\frac{1}{2}XY = n$. Thus, algebraically speaking, the condition that n be a congruent number says that these two equations have a simultaneous solution $X, Y, Z \in \mathbb{Q}$. In the proposition that follows, we derive an alternate condition for n to be a congruent number. In listing triangles with sides X, Y, Z, we shall not want to list X, Y, Z and Y, X, Z separately. So for now let us fix the ordering by requiring that $X < Y < Z$ (Z is the hypotenuse).

Proposition 1. *Let n be a fixed squarefree positive integer. Let X, Y, Z, x always denote positive rational numbers, with $X < Y < Z$. There is a one-to-one correspondence between right triangles with legs X and Y, hypotenuse Z, and area n; and numbers x for which x, $x + n$, and $x - n$ are each the square of a rational number. The correspondence is*:

$$X, Y, Z \to x = (Z/2)^2$$

$$x \to X = \sqrt{x+n} - \sqrt{x-n}, \quad Y = \sqrt{x+n} + \sqrt{x-n}, \quad Z = 2\sqrt{x}.$$

In particular, n is a congruent number if and only if there exists x such that x, $x + n$, and $x - n$ are squares of rational numbers.

Proof. First suppose that X, Y, Z is a triple with the desired properties: $X^2 + Y^2 = Z^2$, $\frac{1}{2}XY = n$. If we add or subtract four times the second equation from the first, we obtain: $(X \pm Y)^2 = Z^2 \pm 4n$. If we then divide both sides by four, we see that $x = (Z/2)^2$ has the property that the numbers $x \pm n$ are the squares of $(X \pm Y)/2$. Conversely, given x with the desired properties, it is easy to see that the three positive rational numbers $X < Y < Z$ given by the formulas in the proposition satisfy: $XY = 2n$, and $X^2 + Y^2 = 4x = Z^2$. Finally, to establish the one-to-one correspondence, it only remains

to verify that no two distinct triples X, Y, Z can lead to the same x. We leave this to the reader (see the problems below). □

PROBLEMS

1. Recall that a Pythagorean triple is a solution (X, Y, Z) in positive integers to the equation $X^2 + Y^2 = Z^2$. It is called "primitive" if X, Y, Z have no common factor. Suppose that $a > b$ are two relatively prime positive integers, not both odd. Show that $X = a^2 - b^2$, $Y = 2ab$, $Z = a^2 + b^2$ form a primitive Pythagorean triple, and that all primitive Pythagorean triples are obtained in this way.

2. Use Problem 1 to write a flowchart for an algorithm that lists all squarefree congruent numbers (of course, not in increasing order). List the first twelve distinct congruent numbers your algorithm gives. Note that there is no way of knowing when a given congruent number n will appear in the list. For example, 101 is a congruent number, but the first Pythagorean triple which leads to an area s^2 101 involves twenty-two-digit numbers (see [Guy 1981, p. 106]). One hundred fifty-seven is even worse (see Fig. I.3). One cannot use this algorithm to establish that some n is *not* a congruent number. Technically, it is not a real algorithm, only a "semi-algorithm".

3. (a) Show that if 1 were a congruent number, then the equation $x^4 - y^4 = u^2$ would have an integer solution with u odd.
   (b) Prove that 1 is not a congruent number. (Note: A consequence is Fermat's Last Theorem for the exponent 4.)

4. Finish the proof of Proposition 1 by showing that no two triples X, Y, Z can lead to the same x.

5. (a) Find $x \in (\mathbb{Q}^+)^2$ such that $x \pm 5 \in (\mathbb{Q}^+)^2$.
   (b) Find $x \in (\mathbb{Q}^+)^2$ such that $x \pm 6 \in (\mathbb{Q}^+)^2$.

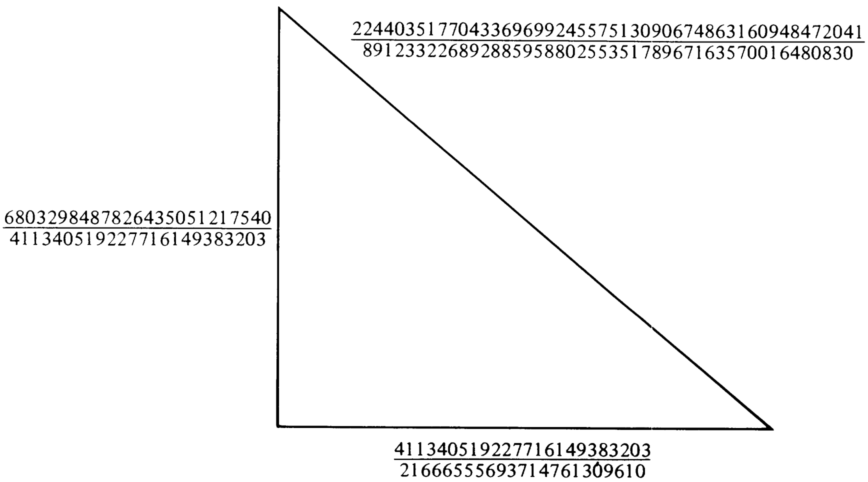


Figure I.3. The Simplest Rational Right Triangle with Area 157 (computed by D. Zagier).

(c) Find two values $x \in (\mathbb{Q}^+)^2$ such that $x \pm 210 \in (\mathbb{Q}^+)^2$. At the end of this chapter we shall prove that if there is one such x, then there are infinitely many. Equivalently (by Proposition 1), if there exists one right triangle with rational sides and area n, then there exist infinitely many.

6. (a) Show that condition (B) in Tunnell's theorem is equivalent to the condition that the number of ways n can be written in the form $2x^2 + y^2 + 8z^2$ with x, y, z integers and z odd, be equal to the number of ways n can be written in this form with z even.
(b) Write a flowchart for an algorithm that tests condition (B) in Tunnell's theorem for a given n.

7. (a) Prove that condition (B) in Tunnell's theorem always holds if n is congruent to 5 or 7 modulo 8.
(b) Check condition (B) for all squarefree $n \equiv 1$ or 3 (mod 8) until you find such an n for which condition (B) holds.
(c) By Tunnell's theorem, the number you found in part (b) should be the smallest congruent number congruent to 1 or 3 modulo 8. Use the algorithm in Problem 2 to find a right triangle with rational sides and area equal to the number you found in part (b).

§2. A certain cubic equation

In this section we find yet another equivalent characterization of congruent numbers.

In the proof of Proposition 1 in the last section, we arrived at the equations $((X \pm Y)/2)^2 = (Z/2)^2 \pm n$ whenever X, Y, Z are the sides of a triangle with area n. If we multiply together these two equations, we obtain $((X^2 - Y^2)/4)^2 = (Z/2)^4 - n^2$. This shows that the equation $u^4 - n^2 = v^2$ has a rational solution, namely, $u = Z/2$ and $v = (X^2 - Y^2)/4$. We next multiply through by u^2 to obtain $u^6 - n^2u^2 = (uv)^2$. If we set $x = u^2 = (Z/2)^2$ (this is the same x as in Proposition 1) and further set $y = uv = (X^2 - Y^2)Z/8$, then we have a pair of rational numbers (x, y) satisfying the cubic equation:

$$y^2 = x^3 - n^2x.$$

Thus, given a right triangle with rational sides X, Y, Z and area n, we obtain a point (x, y) in the xy-plane having rational coordinates and lying on the curve $y^2 = x^3 - n^2x$. Conversely, can we say that any point (x, y) with $x, y \in \mathbb{Q}$ which lies on the cubic curve must necessarily come from such a right triangle? Obviously not, because in the first place the x-coordinate $x = u^2 = (Z/2)^2$ must lie in $(\mathbb{Q}^+)^2$ if the point (x, y) can be obtained as in the last paragraph. In the second place, we can see that the x-coordinate of such a point must have its denominator divisible by 2. To see this, notice that the triangle X, Y, Z can be obtained starting with a primitive Pythagorean triple X', Y', Z' corresponding to a right triangle with integral sides X', Y', Z' and area s^2n, and then dividing the sides by s to get X, Y, Z. But in a primitive

Pythagorean triple X' and Y' have different parity, and Z' is odd. We conclude that (1) $x = (Z/2)^2 = (Z'/2s)^2$ has denominator divisible by 2 and (2) the power of 2 dividing the denominator of Z is equal to the power of 2 dividing the denominator of one of the other two sides, while a strictly lower power of 2 divides the denominator of the third side. (For example, in the triangle in Fig. I.2 with area 5, the hypotenuse and the shorter side have a 2 in the denominator, while the other leg does not.) We conclude that a *necessary* condition for the point (x, y) with rational coordinates on the curve $y^2 = x^3 - n^2x$ to come from a right triangle is that x be a square and that its denominator be divisible by 2. For example, when $n = 31$, the point $(41^2/7^2, 29520/7^3)$ on the curve $y^2 = x^3 - 31^2x$ does not come from a triangle, even though its x-coordinate is a square.

Finally, a third necessary condition is that the numerator of x have no common factor with n. To see this, suppose that $p > 2$ is a prime dividing both n and the numerator of x. Then p divides the numerator of $x \pm n = ((X \pm Y)/2)^2$, and so it also divides the numerators of $(X + Y)/2$ and $(X - Y)/2$. Then p divides the numerators of the sum X and the difference Y. Hence p^2 divides $n = \frac{1}{2}XY$. But n was assumed to be squarefree. This contradiction shows that x must be a square with even denominator and numerator prime to n. A numerical example (for which I thank Clas Löfwall) showing that the first two conditions alone are not sufficient is provided by the point $(x, y) = (25/4, 75/8)$ on the curve $y^2 = x^3 - n^2x$, $n = 5$.

We next prove that these three conditions are not only necessary but also *sufficient* for a point on the curve to come from a triangle.

Proposition 2. *Let (x, y) be a point with rational coordinates on the curve $y^2 = x^3 - n^2x$. Suppose that x satisfies the three conditions:* (i) *it is the square of a rational number,* (ii) *its denominator is even, and* (iii) *its numerator has no common factor with n. Then there exists a right triangle with rational sides and area n which corresponds to x under the correspondence in Proposition* 1.

PROOF. Let $u = \sqrt{x} \in \mathbb{Q}^+$. We work backwards through the sequence of steps at the beginning of this section. That is, set $v = y/u$, so that $v^2 = y^2/x = x^2 - n^2$, i.e., $v^2 + n^2 = x^2$. Now let t be the denominator of u, i.e., the smallest positive integer such that $tu \in \mathbb{Z}$. By assumption, t is even. Notice that the denominators of v^2 and x^2 are the same (because n is an integer, and $v^2 + n^2 = x^2$), and this denominator is t^4. Thus, t^2v, t^2n, t^2x is a primitive Pythagorean triple, with t^2n even. (Here the primitivity of the triple follows from condition (iii).) By Problem 1 of §1, there exist integers a and b such that: $t^2n = 2ab$, $t^2v = a^2 - b^2$, $t^2x = a^2 + b^2$. Then the right triangle with sides $2a/t$, $2b/t$, $2u$ has area $2ab/t^2 = n$, as desired. The image of this triangle $X = 2a/t$, $Y = 2b/t$, $Z = 2u$ under the correspondence in Proposition 1 is $x = (Z/2)^2 = u^2$. This proves Proposition 2. □

We shall later prove another characterization of the points $P = (x, y)$ on the curve $y^2 = x^3 - n^2x$ which correspond to rational right triangles of

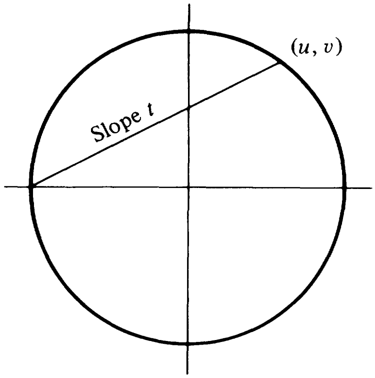


Figure I.4

area n. Namely, they are the points $P = (x, y)$ which are "twice" a rational point $P' = (x', y')$. That is, $P' + P' = P$, where "+" is an addition law for points on our curve, which we shall define later.

Problems

1. Find a simple linear change of variables that gives a one-to-one correspondence between points on $ny^2 = x^3 + ax^2 + bx + c$ and points on $y^2 = x^3 + anx^2 + bn^2x + cn^3$. For example, an alternate form of the equation $y^2 = x^3 - n^2x$ is the equation $ny^2 = x^3 - x$.

2. Another correspondence between rational right triangles X, Y, Z with area $\frac{1}{2}XY = n$ and rational solutions to $y^2 = x^3 - n^2x$ can be constructed as follows.

   (a) Parametrize all right triangles by letting the point $u = X/Z$, $v = Y/Z$ on the unit circle correspond to the slope t of the line joining $(-1, 0)$ to this point (see Fig. I.4). Show that
$$u = \frac{1 - t^2}{1 + t^2}, \qquad v = \frac{2t}{1 + t^2}.$$
   (Note: This is the usual way to parametrize a conic. If $t = a/b$ is rational, then the point (u, v) corresponds to the Pythagorean triple constructed by the method at the beginning of the chapter.)

   (b) If we want the triangle X, Y, Z to have area n, express n/Z^2 in terms of t.

   (c) Show that the point $x = -nt$, $y = n^2(1 + t^2)/Z$ is on the curve $y^2 = x^3 - n^2x$. Express (x, y) in terms of X, Y, Z.

   (d) Conversely, show that any point (x, y) on the curve $y^2 = x^3 - n^2x$ with $y \neq 0$ comes from a triangle, except that to get points with positive x, we must allow triangles with negative X and Y (but positive area $\frac{1}{2}XY = n$), and to get points with negative y we must allow negative Z (see Fig. I.5). Later in this chapter we shall show the connection between this correspondence and the one given in the text above.

   (e) Find the points on $y^2 = x^3 - 36x$ coming from the 3–4–5 right triangle and all equivalent triangles (4–3–5, (-3)–(-4)–5, etc.).

3. Generalize the congruent number problem as follows. Fix an angle θ not necessarily 90°. But suppose that $A = \cos\theta$ and $B = \sin\theta$ are both rational. Let n be a square-

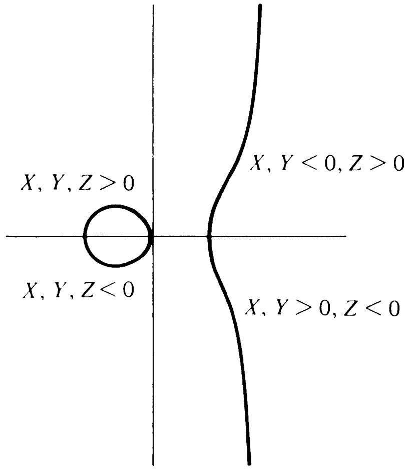


Figure I.5

free natural number. One can then ask whether n is the area of any triangle with rational sides one of whose angles is θ.

(a) Show that the answer to this question is equivalent to a question about rational solutions to a certain cubic equation (whose coefficients depend on θ as well as n).

(b) Suppose that the line joining the point $(-1, 0)$ to the point (A, B) on the unit circle has slope λ. Show that the cubic in part (a) is equivalent (by a linear change of variables) to the cubic $ny^2 = x(x - \lambda)(x + (1/\lambda))$. The classical congruent number problem is, of course, the case $\lambda = 1$.

§3. Elliptic curves

The locus of points $P = (x, y)$ satisfying $y^2 = x^3 - n^2x$ is a special case of what's called an "elliptic curve". More generally, let K be any field, and let $f(x) \in K[x]$ be a cubic polynomial with coefficients in K which has *distinct* roots (perhaps in some extension of K). We shall suppose that K does *not* have characteristic 2. Then the solutions to the equation

$$y^2 = f(x), \tag{3.1}$$

where x and y are in some extension K' of K, are called the *K'-points of the elliptic curve defined by* (3.1). We have just been dealing with the example $K = K' = \mathbb{Q}$, $f(x) = x^3 - n^2x$. Note that this example $y^2 = x^3 - n^2x$ satisfies the condition for an elliptic curve over any field K of characteristic p, as long as p does not divide $2n$, since the three roots $0, \pm n$ of $f(x) = x^3 - n^2x$ are then distinct.

In general, if $x_0, y_0 \in K'$ are the coordinates of a point on a curve C defined by an equation $F(x, y) = 0$, we say that C is "smooth" at (x_0, y_0) if the two partial derivatives $\partial F/\partial x$ and $\partial F/\partial y$ are not both zero at (x_0, y_0).

This is the definition regardless of the ground field (the partial derivative of a polynomial $F(x, y)$ is defined by the usual formula, which makes sense over any field). If K' is the field $\mathbb{R}$ of real numbers, this agrees with the usual condition for C to have a tangent line. In the case $F(x, y) = y^2 - f(x)$, the partial derivatives are $2y_0$ and $-f'(x_0)$. Since K' is not a field of characteristic 2, these vanish simultaneously if and only if $y_0 = 0$ and x_0 is a multiple root of $f(x)$. Thus, the curve has a non-smooth point if and only if $f(x)$ has a multiple root. It is for this reason that we assumed distinct roots in the definition of an elliptic curve: an elliptic curve is smooth at all of its points.

In addition to the points (x, y) on an elliptic curve (3.1), there is a very important "point at infinity" that we would like to consider as being on the curve, much as in complex variable theory in addition to the points on the complex plane one throws in a point at infinity, thereby forming the "Riemann sphere". To do this precisely, we now introduce projective coordinates.

By the "total degree" of a monomial $x^i y^j$ we mean $i + j$. By the "total degree" of a polynomial $F(x, y)$ we mean the maximum total degree of the monomials that occur with nonzero coefficients. If $F(x, y)$ has total degree n, we define the corresponding *homogeneous polynomial* $\tilde{F}(x, y, z)$ of *three* variables to be what you get by multiplying each monomial $x^i y^j$ in $F(x, y)$ by z^{n-i-j} to bring its total degree in the variables x, y, z up to n; in other words,

$$\tilde{F}(x, y, z) = z^n F\left(\frac{x}{z}, \frac{y}{z}\right).$$

In our example $F(x, y) = y^2 - (x^3 - n^2 x)$, we have $\tilde{F}(x, y, z) = y^2 z - x^3 + n^2 x z^2$. Notice that $F(x, y) = \tilde{F}(x, y, 1)$.

Suppose that our polynomials have coefficients in a field K, and we are interested in triples $x, y, z \in K$ such that $\tilde{F}(x, y, z) = 0$. Notice that:

(1) for any $\lambda \in K$, $\tilde{F}(\lambda x, \lambda y, \lambda z) = \lambda^n \tilde{F}(x, y, z)$ (n = total degree of F);
(2) for any nonzero $\lambda \in K$, $\tilde{F}(\lambda x, \lambda y, \lambda z) = 0$ if and only if $\tilde{F}(x, y, z) = 0$. In particular, for $z \neq 0$ we have $\tilde{F}(x, y, z) = 0$ if and only if $F(x/z, y/z) = 0$.

Because of (2), it is natural to look at equivalence classes of triples $x, y, z \in K$, where we say that two triples (x, y, z) and (x', y', z') are equivalent if there exists a nonzero $\lambda \in K$ such that $(x', y', z') = \lambda(x, y, z)$. We omit the trivial triple $(0, 0, 0)$, and then we define the "projective plane $\mathbb{P}_K^2$" to be the set of all equivalence classes of nontrivial triples.

No normal person likes to think in terms of "equivalence classes", and fortunately there are more visual ways to think of the projective plane. Suppose that K is the field $\mathbb{R}$ of real numbers. Then the triples (x, y, z) in an equivalence class all correspond to points in three-dimensional Euclidean space lying on a line through the origin. Thus, $\mathbb{P}_{\mathbb{R}}^2$ can be thought of geometrically as the set of lines through the origin in three-dimensional space.

Another way to visualize $\mathbb{P}_{\mathbb{R}}^2$ is to place a plane at a distance from the origin in three-dimensional space, for example, take the plane parallel to the xy-plane and at a distance 1 from it, i.e., the plane with equation $z = 1$. All

lines through the origin, except for those lying in the xy-plane, have a unique point of intersection with this plane. That is, every equivalence class of triples (x, y, z) with nonzero z-coordinate has a unique triple of the form $(x, y, 1)$. So we think of such equivalence classes as points in the ordinary xy-plane. The remaining triples, those of the form $(x, y, 0)$, make up the "line at infinity".

The line at infinity, in turn, can be visualized as an ordinary line (say, the line $y = 1$ in the xy-plane) consisting of the equivalence classes with nonzero y-coordinate and hence containing a unique triple of the form $(x, 1, 0)$, together with a single "point at infinity" $(1, 0, 0)$. That is, we define the projective line $\mathbb{P}_K^1$ over a field K to be the set of equivalence classes of pairs (x, y) with $(x, y) \sim (\lambda x, \lambda y)$. Then $\mathbb{P}_K^2$ can be thought of as an ordinary plane $(x, y, 1)$ together with a projective line at infinity, which, in turn, consists of an ordinary line $(x, 1, 0)$ together with its point at infinity $(1, 0, 0)$.

More generally, n-dimensional projective space $\mathbb{P}_K^n$ is defined using equivalence classes of $(n + 1)$-tuples, and can be visualized as the usual space of n-tuples $(x_1, \ldots, x_n, 1)$ together with a $\mathbb{P}_K^{n-1}$ at infinity. But we shall only have need of $\mathbb{P}_K^1$ and $\mathbb{P}_K^2$.

Given a homogeneous polynomial $\tilde{F}(x, y, z)$ with coefficients in K, we can look at the solution set consisting of points (x, y, z) in $\mathbb{P}_K^2$ (actually, equivalence classes of (x, y, z)) for which $\tilde{F}(x, y, z) = 0$. The points of this solution set where $z \neq 0$ are the points $(x, y, 1)$ for which $\tilde{F}(x, y, 1) = F(x, y) = 0$. The remaining points are on the line at infinity. The solution set of $\tilde{F}(x, y, z) = 0$ is called the "projective completion" of the curve $F(x, y) = 0$. From now on, when we speak of a "line", a "conic section", an "elliptic curve", etc., we shall usually be working in a projective plane $\mathbb{P}_K^2$, in which case these terms will always denote the projective completion of the usual curve in the xy-plane. For example, the line $y = mx + b$ will really mean the solution set to $y = mx + bz$ in $\mathbb{P}_K^2$; and the elliptic curve $y^2 = x^3 - n^2x$ will now mean the solution set to $y^2z = x^3 - n^2xz^2$ in $\mathbb{P}_K^2$.

Let us look more closely at our favorite example: $F(x, y) = y^2 - x^3 + n^2x$, $\tilde{F}(x, y, z) = y^2z - x^3 + n^2xz^2$. The points at infinity on this elliptic curve are the equivalence classes $(x, y, 0)$ such that $0 = \tilde{F}(x, y, 0) = -x^3$. i.e., $x = 0$. There is only one such equivalence class $(0, 1, 0)$. Intuitively, if we take $K = \mathbb{R}$, we can think of the curve $y^2 = x^3 - n^2x$ heading off in an increasingly vertical direction as it approaches the line at infinity (see Fig. I.6). The points on the line at infinity correspond to the lines through the origin in the xy-plane, i.e., there is one for every possible slope y/x of such a line. As we move far out along our elliptic curve, we approach slope $y/x = \infty$, corresponding to the single point $(0, 1, 0)$ on the line at infinity. Notice that any elliptic curve $y^2 = f(x)$ similarly contains exactly one point at infinity $(0, 1, 0)$.

All of the usual concepts of calculus on curves $F(x, y) = 0$ in the xy-plane carry over to the corresponding projective curve $\tilde{F}(x, y, z) = 0$. Such notions as the tangent line at a point, points of inflection, smooth and singular points all depend only upon what is happening in a neighborhood of the

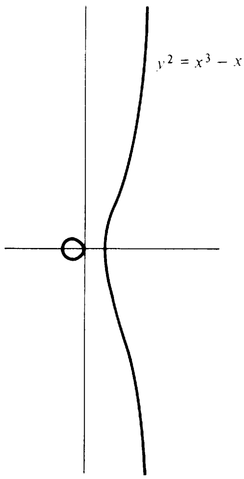


Figure I.6

point in question. And any point in $\mathbb{P}^2_{\mathbb{R}}$ has a large neighborhood which looks like an ordinary plane. More precisely, if we are interested in a point with nonzero z-coordinate, we can work in the usual xy-plane, where the curve has equation $F(x, y) = \tilde{F}(x, y, 1) = 0$. If we want to examine a point on the line $z = 0$, however, we put the triple in either the form $(x, 1, 0)$ or $(1, y, 0)$. In the former case, we think of it as a point on the curve $F(x, 1, z) = 0$ in the xz-plane; and in the latter case as a point on the curve $F(1, y, z) = 0$ in the yz-plane.

For example, near the point at infinity $(0, 1, 0)$ on the elliptic curve $y^2z - x^3 + n^2xz^2 = 0$, all points have the form $(x, 1, z)$ with $z - x^3 + n^2xz^2 = 0$. The latter equation, in fact, gives us all points on the elliptic curve except for the three points $(0, 0, 1)$, $(\pm n, 0, 1)$ having zero y-coordinate (these are the three "points at infinity" if we think in terms of xz-coordinates).

Problems

1. Prove that if K is an infinite field and $F(x, y, z) \in K[x, y, z]$ satisfies $F(\lambda x, \lambda y, \lambda z) = \lambda^n F(x, y, z)$ for all $\lambda, x, y, z \in K$, then F is homogeneous, i.e., each monomial has total degree n. Give a counterexample if K is finite.

2. By a "line" in $\mathbb{P}^2_K$ we mean either the projective completion of a line in the xy-plane or the line at infinity. Show that a line in $\mathbb{P}^2_K$ has equation of the form $ax + by + cz = 0$, with $a, b, c \in K$ not all zero; and that two such equations determine the same line if and only if the two triples (a, b, c) differ by a multiple. Construct a 1-to-1 correspondence between lines in a copy of $\mathbb{P}^2_K$ with coordinates (x, y, z) and points in another copy of $\mathbb{P}^2_K$ with coordinates (a, b, c) and between points in the xyz-projective plane and lines in the abc-projective plane, such that a bunch of points are on the same line in the first projective plane if and only if the lines that correspond to them in the second projective plane all meet in the same point. The xyz-projective plane and the abc-projective plane are called the "duals" of each other.

3. How many points at infinity are on a parabola in $\mathbb{P}^2_{\mathbb{R}}$? an ellipse? a hyperbola?

4. Prove that any two nondegenerate conic sections in $\mathbb{P}^2_{\mathbb{R}}$ are equivalent to one another by some linear change of variables.

5. (a) If $\tilde{F}(x, y, z) \in K[x, y, z]$ is homogeneous of degree n, show that
$$x\frac{\partial \tilde{F}}{\partial x} + y\frac{\partial \tilde{F}}{\partial y} + z\frac{\partial \tilde{F}}{\partial z} = n\tilde{F}.$$
(b) If K has characteristic zero, show that a point $(x, y, z) \in \mathbb{P}^2_K$ is a non-smooth point on the curve C: $\tilde{F}(x, y, z) = 0$ if and only if the triple $(\partial\tilde{F}/\partial x, \partial\tilde{F}/\partial y, \partial\tilde{F}/\partial z)$ is $(0, 0, 0)$ at our particular (x, y, z). Give a counterexample if char $K \neq 0$. In what follows, suppose that char $K = 0$, e.g., $K = \mathbb{R}$.
(c) Show that the tangent line to C at a smooth point (x_0, y_0, z_0) has equation $ax + by + cz = 0$, where
$$a = \left.\frac{\partial \tilde{F}}{\partial x}\right|_{(x_0, y_0, z_0)}, \qquad b = \left.\frac{\partial \tilde{F}}{\partial y}\right|_{(x_0, y_0, z_0)}, \qquad c = \left.\frac{\partial \tilde{F}}{\partial z}\right|_{(x_0, y_0, z_0)}.$$
(d) Prove that the condition that (x, y, z) be a smooth point on C does not depend upon the choice of coordinates, i.e., it does not change if we shift to $x'y'z'$-coordinates, where $(x' \ y' \ z') = (x \ y \ z)A$ with A an invertible 3×3 matrix. For example, if more than one of the coordinates are nonzero, it makes no difference which we choose to regard as the "z-coordinate", i.e., whether we look at C in the xy-plane, the xz-plane, or the yz-plane.
(e) Prove that the condition that a given line l be tangent to C at a smooth point (x, y, z) does not depend upon the choice of coordinates.

6. (a) Let $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$ be two distinct points in $\mathbb{P}^2_K$. Show that the line joining P_1 and P_2 can be given in parametrized form as $sP_1 + tP_2$, i.e., $\{(sx_1 + tx_2, sy_1 + ty_2, sz_1 + tz_2) | s, t \in K\}$. Check that this linear map takes $\mathbb{P}^1_K$ (with coordinates s, t) bijectively onto the line $\overline{P_1P_2}$ in $\mathbb{P}^2_K$. What part of the line do you get by taking $s = 1$ and letting t vary?
(b) Suppose that $K = \mathbb{R}$ or $\mathbb{C}$. If the curve $F(x, y) = 0$ in the xy-plane is smooth at $P_1 = (x_1, y_1)$ with nonvertical tangent line, then we can expand the implicit function $y = f(x)$ in a Taylor series about $x = x_1$. The linear term gives the tangent line. If we subtract off the linear term, we obtain $f(x) - y_1 - f'(x_1)(x - x_1) = a_m(x - x_1)^m + \cdots$, where $a_m \neq 0, m \geq 2$. m is called the "order of tangency". We say that (x_1, y_1) is a point of inflection if $m > 2$, i.e., $f''(x_1) = 0$. (In the case $K = \mathbb{R}$, note that we are not requiring a change in concavity with this definition, e.g., $y = x^4$ has a point of inflection at $x = 0$.) Let $P_1 = (x_1, y_1, z_1)$, $z_1 \neq 0$, and let $l = \overline{P_1P_2}$ be tangent to the curve $F(x, y) = \tilde{F}(x, y, 1)$ at the smooth point P_1. Let $P_2 = (x_2, y_2, z_2)$. Show that m is the lowest power of t that occurs in $\tilde{F}(x_1 + tx_2, y_1 + ty_2, z_1 + tz_2) \in K[t]$.
(c) Show that m does not change if we make a linear change of variables in $\mathbb{P}^2_K$. For example, suppose that y_1 and z_1 are both nonzero, and we use the xz-plane instead of the xy-plane in parts (a) and (b).

7. Show that the line at infinity (with equation $z = 0$) is tangent to the elliptic curve $y^2 = f(x)$ at $(0, 1, 0)$, and that the point $(0, 1, 0)$ is a point of inflection on the curve.

§4. Doubly periodic functions

Let L be a lattice in the complex plane, by which we mean the set of all integral linear combinations of two given complex numbers ω_1 and ω_2, where ω_1 and ω_2 do not lie on the same line through the origin. For example, if $\omega_1 = i$ and $\omega_2 = 1$, we get the lattice of Gaussian integers $\{mi + n | m, n \in \mathbb{Z}\}$. It will turn out that the example of the lattice of Gaussian integers is intimately related to the elliptic curves $y^2 = x^3 - n^2x$ that come from the congruent number problem.

The fundamental parallelogram for ω_1, ω_2 is defined as

$$\Pi = \{a\omega_1 + b\omega_2 | 0 \leq a \leq 1, 0 \leq b \leq 1\}.$$

Since ω_1, ω_2 form a basis for $\mathbb{C}$ over $\mathbb{R}$, any number $x \in \mathbb{C}$ can be written in the form $x = a\omega_1 + b\omega_2$ for some $a, b \in \mathbb{R}$. Then x can be written as the sum of an element in the lattice $L = \{m\omega_1 + n\omega_2\}$ and an element in Π, and in only one way unless a or b happens to be an integer, i.e., the element of Π happens to lie on the boundary $\partial\Pi$.

We shall always take ω_1, ω_2 in clockwise order; that is, we shall assume that ω_1/ω_2 has positive imaginary part.

Notice that the choice of ω_1, ω_2 giving the lattice L is not unique. For example, $\omega_1' = \omega_1 + \omega_2$ and ω_2 give the same lattice. More generally, we can obtain new bases ω_1', ω_2' for the lattice L by applying a matrix with integer entries and determinant 1 (see Problem 1 below).

For a given lattice L, a meromorphic function on $\mathbb{C}$ is said to be an *elliptic function* relative to L if $f(z + l) = f(z)$ for all $l \in L$. Notice that it suffices to check this property for $l = \omega_1$ and $l = \omega_2$. In other words, an elliptic function is periodic with two periods ω_1 and ω_2. Such a function is determined by its values on the fundamental parallelogram Π; and its values on opposite points of the boundary of Π are the same, i.e., $f(a\omega_1 + \omega_2) = f(a\omega_1)$, $f(\omega_1 + b\omega_2) = f(b\omega_2)$. Thus, we can think of an elliptic function $f(z)$ as a function on the set Π with opposite sides glued together. This set (more precisely, "complex manifold") is known as a "torus". It looks like a donut.

Doubly periodic functions on the complex numbers are directly analogous to singly periodic functions on the real numbers. A function $f(x)$ on $\mathbb{R}$ which satisfies $f(x + n\omega) = f(x)$ is determined by its values on the interval $[0, \omega]$. Its values at 0 and ω are the same, so it can be thought of as a function on the interval $[0, \omega]$ with the endpoints glued together. The "real manifold" obtained by gluing the endpoints is simply a circle (see Fig. I.7).

Returning now to elliptic functions for a lattice L, we let $\mathscr{E}_L$ denote the set of such functions. We immediately see that $\mathscr{E}_L$ is a subfield of the field of all meromorphic functions, i.e., the sum, difference, product, or quotient of two elliptic functions is elliptic. In addition, the subfield $\mathscr{E}_L$ is closed under differentiation. We now prove a sequence of propositions giving some very special properties which any elliptic function must have. The condition that a meromorphic function be doubly periodic turns out to be much more

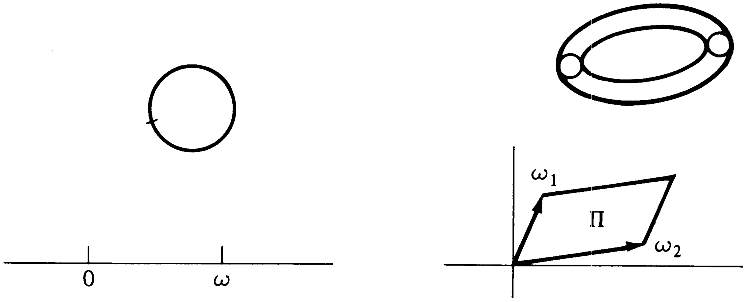


Figure I.7

restrictive than the analogous condition in the real case. The set of real-analytic periodic functions with given period is much "larger" than the set $\mathscr{E}_L$ of elliptic functions for a given period lattice L.

Proposition 3. *A function $f(z) \in \mathscr{E}_L$, $L = \{m\omega_1 + n\omega_2\}$, which has no pole in the fundamental parallelogram Π must be a constant.*

PROOF. Since Π is compact, any such function must be bounded on Π, say by a constant M. But then $|f(z)| < M$ for all z, since the values of $f(z)$ are determined by the values on Π. By Liouville's theorem, a meromorphic function which is bounded on all of $\mathbb{C}$ must be a constant. □

Proposition 4. *With the same notation as above, let $\alpha + \Pi$ denote the translate of Π by the complex number α, i.e., $\{\alpha + z | z \in \Pi\}$. Suppose that $f(z) \in \mathscr{E}_L$ has no poles on the boundary C of $\alpha + \Pi$. Then the sum of the residues of $f(z)$ in $\alpha + \Pi$ is zero.*

PROOF. By the residue theorem, this sum is equal to

$$\frac{1}{2\pi i}\int_C f(z)dz.$$

But the integral over opposite sides cancel, since the values of $f(z)$ at corresponding points are the same, while dz has opposite signs, because the path of integration is in opposite directions on opposite sides (see Fig. I.8). Thus, the integral is zero, and so the sum of residues is zero. □

Notice that, since a meromorphic function can only have finitely many poles in a bounded region, it is always possible to choose an α such that the boundary of $\alpha + \Pi$ misses the poles of $f(z)$. Also note that Proposition 4 immediately implies that a nonconstant $f(z) \in \mathscr{E}_L$ must have at least two poles (or a multiple pole), since if it had a single simple pole, then the sum of residues would not be zero.

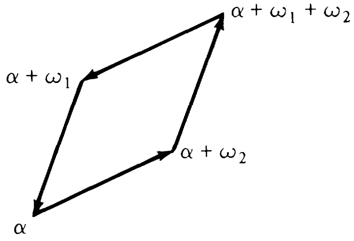


Figure I.8

Proposition 5. *Under the conditions of Proposition* 4, *suppose that* $f(z)$ *has no zeros or poles on the boundary of* $\alpha + \Pi$. *Let* $\{m_i\}$ *be the orders of the various zeros in* $\alpha + \Pi$, *and let* $\{n_j\}$ *be the orders of the various poles. Then* $\Sigma\, m_i = \Sigma\, n_j$.

PROOF. Apply Proposition 4 to the elliptic function $f'(z)/f(z)$. Recall that the logarithmic derivative $f'(z)/f(z)$ has a pole precisely where $f(z)$ has a zero or pole, such a pole is simple, and the residue there is equal to the order of zero or pole of the original $f(z)$ (negative if a pole). (Recall the argument: If $f(z) = c_m(z-a)^m + \cdots$, then $f'(z) = c_m m(z-a)^{m-1} + \cdots$, and so $f'(z)/f(z) = m(z-a)^{-1} + \cdots$.) Thus, the sum of the residues of $f'(z)/f(z)$ is $\Sigma\, m_i - \Sigma\, n_j = 0$. □

We now define what will turn out to be a key example of an elliptic function relative to the lattice $L = \{m\omega_1 + n\omega_2\}$. This function is called the Weierstrass $\wp$-function. It is denoted $\wp(z; L)$ or $\wp(z; \omega_1, \omega_2)$, or simply $\wp(z)$ if the lattice is fixed throughout the discussion. We set

$$\wp(z) = \wp(z; L) \underset{\text{def}}{=} \frac{1}{z^2} + \sum_{\substack{l \in L \\ l \neq 0}} \left(\frac{1}{(z-l)^2} - \frac{1}{l^2} \right). \tag{4.1}$$

Proposition 6. *The sum in* (4.1) *converges absolutely and uniformly for z in any compact subset of* $\mathbb{C} - L$.

PROOF. The sum in question is taken over a two-dimensional lattice. The proof of convergence will be rather routine if we keep in mind a one-dimensional analog. If instead of L we take the integers $\mathbb{Z}$, and instead of reciprocal squares we take reciprocals, we obtain a real function $f(x) = \frac{1}{x} + \Sigma \frac{1}{x-l} + \frac{1}{l}$, where the sum is over nonzero $l \in \mathbb{Z}$. To prove absolute and uniform convergence in any compact subset of $\mathbb{R} - \mathbb{Z}$, first write the summand as $x/(l(x-l))$, and then use a comparison test, showing that the series in question basically has the same behavior as l^{-2}. More precisely, use the following lemma: if $\Sigma\, b_l$ is a convergent sum of positive terms (all our sums being over nonzero $l \in \mathbb{Z}$), and if $\Sigma f_l(x)$ has the property that $|f_l(x)/b_l|$ approaches a finite limit as $l \to \pm\infty$, uniformly for x in some set, then the sum $\Sigma f_l(x)$ converges absolutely and uniformly for x in that set. The details

are easy to fill in. (By the way, our particular example of $f(x)$ can be shown to be the function $\pi \cot \pi x$; just take the logarithmic derivative of both sides of the infinite product for the sine function: $\sin \pi x = \pi x \Pi_{n=1}^{\infty}(1 - x^2/n^2)$.)

The proof of Proposition 6 proceeds in the same way. First write the summand over a common denominator:

$$\frac{1}{(z-l)^2} - \frac{1}{l^2} = \frac{2z - z^2/l}{(z-l)^2 l}.$$

Then show absolute and uniform convergence by comparison with $|l|^{-3}$, where the sum is taken over all nonzero $l \in L$. More precisely, Proposition 6 will follow from the following two lemmas.

Lemma 1. *If $\Sigma\, b_l$ is a convergent sum of positive terms, where the sum is taken over all nonzero elements in the lattice L, and if $\Sigma\, f_l(z)$ has the property that $|f_l(z)/b_l|$ approaches a finite limit as $|l| \to \infty$, uniformly for z in some subset of $\mathbb{C}$, then the sum $\Sigma\, f_l(z)$ converges absolutely and uniformly for z in that set.*

Lemma 2. *$\Sigma\, |l|^{-s}$ converges if $s > 2$.*

The proof of Lemma 1 is routine, and will be omitted. We give a sketch of the proof of Lemma 2. We split the sum into sums over l satisfying $n - 1 < |l| \leq n$, as $n = 1, 2, \ldots$. It is not hard to show that the number of l in that annulus has order of magnitude n. Thus, the sum in the lemma is bounded by a constant times $\Sigma_{n=1}^{\infty} n \cdot n^{-s} = \Sigma\, n^{1-s}$, and the latter sum converges for $s - 1 > 1$.

This concludes the proof of Proposition 6. □

Proposition 7. *$\wp(z) \in \mathscr{E}_L$, and its only pole is a double pole at each lattice point.*

Proof. The same argument as in the proof of Proposition 6 shows that for any fixed $l \in L$, the function $\wp(z) - (z - l)^{-2}$ is continuous at $z = l$. Thus, $\wp(z)$ is a meromorphic function with a double pole at all lattice points and no other poles. Next, note that $\wp(z) = \wp(-z)$, because the right side of (4.1) remains unchanged if z is replaced by $-z$ and l is replaced by $-l$; but summing over $l \in L$ is the same as summing over $-l \in L$.

To prove double periodicity, we look at the derivative. Differentiating (4.1) term-by-term, we obtain:

$$\wp'(z) = -2 \sum_{l \in L} \frac{1}{(z-l)^3}.$$

Now $\wp'(z)$ is obviously doubly periodic, since replacing z by $z + l_0$ for some fixed $l_0 \in L$ merely rearranges the terms in the sum. Thus, $\wp'(z) \in \mathscr{E}_L$. To prove that $\wp(z) \in \mathscr{E}_L$, it suffices to show that $\wp(z + \omega_i) - \wp(z) = 0$ for $i = 1, 2$. We prove this for $i = 1$; the identical argument applies to $i = 2$.

Since the derivative of the function $\wp(z+\omega_1)-\wp(z)$ is $\wp'(z+\omega_1)-\wp'(z)=0$, we must have $\wp(z+\omega_1)-\wp(z)=C$ for some constant C. But substituting $z=-\frac{1}{2}\omega_1$ and using the fact that $\wp(z)$ is an even function, we conclude that $C=\wp(\frac{1}{2}\omega_1)-\wp(-\frac{1}{2}\omega_1)=0$. This concludes the proof. □

Notice that the double periodicity of $\wp(z)$ was not immediately obvious from the definition (4.1).

Since $\wp(z)$ has exactly one double pole in a fundamental domain of the form $\alpha+\Pi$, by Proposition 5 it has exactly two zeros there (or one double zero). The same is true of any elliptic function of the form $\wp(z)-u$, where u is a constant. It is not hard to show (see the problems below) that $\wp(z)$ takes every value $u\in\mathbb{C}\cup\{\infty\}$ exactly twice on the torus (i.e., a fundamental parallelogram with opposite sides glued together), counting multiplicity (which means the order of zero of $\wp(z)-u$); and that the values assumed with multiplicity two are ∞, $e_1 \underset{\text{def}}{=} \wp(\omega_1/2)$, $e_2 \underset{\text{def}}{=} \wp(\omega_2/2)$, and $e_3 \underset{\text{def}}{=} \wp((\omega_1+\omega_2)/2)$. Namely, $\wp(z)$ has a double pole at 0, while the other three points are the zeros of $\wp'(z)$.

§5. The field of elliptic functions

Proposition 7 gives us a concrete example of an elliptic function. Just as $\sin x$ and $\cos x$ play a basic role in the theory of periodic functions on $\mathbb{R}$, because of Fourier expansion, similarly the functions $\wp(z)$ and $\wp'(z)$ play a fundamental role in the study of elliptic functions. But unlike in the real case, we do not even need infinite series to express an arbitrary elliptic function in terms of these two basic ones.

Proposition 8. $\mathscr{E}_L=\mathbb{C}(\wp,\wp')$, *i.e., any elliptic function for L is a rational expression in* $\wp(z;L)$ *and* $\wp'(z;L)$. *More precisely, given* $f(z)\in\mathscr{E}_L$, *there exist two rational functions* $g(X)$, $h(X)$ *such that* $f(z)=g(\wp(z))+\wp'(z)h(\wp(z))$.

Proof. If $f(z)$ is an elliptic function for L, then so are the two even functions

$$\frac{f(z)+f(-z)}{2} \quad\text{and}\quad \frac{f(z)-f(-z)}{2\wp'(z)}.$$

Since $f(z)$ is equal to the first of these functions plus $\wp'(z)$ times the second, to prove Proposition 8 it suffices to prove

Proposition 9. *The subfield* $\mathscr{E}_L^+\subset\mathscr{E}_L$ *of even elliptic functions for L is generated by* $\wp(z)$, *i.e.,* $\mathscr{E}_L^+=\mathbb{C}(\wp)$.

Proof. The idea of the proof is to cook up a function which has the same zeros and poles as $f(z)$ using only functions of the form $\wp(z)-u$ with u a constant.

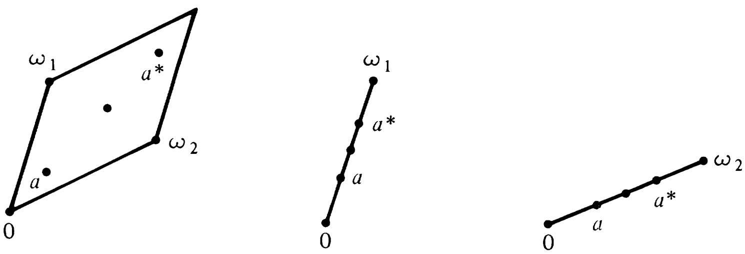


Figure I.9

The ratio of $f(z)$ to such a function is an elliptic function with no poles, and so must be a constant, by Proposition 3.

Let $f(z) \in \mathscr{E}_L^+$. We first list the zeros and poles of $f(z)$. But we must do this carefully, in a special way. Let Π' be a fundamental parallelogram with two sides removed: $\Pi' = \{a\omega_1 + b\omega_2 | 0 \leq a < 1, 0 \leq b < 1\}$. Then every point in $\mathbb{C}$ differs by a lattice element from *exactly* one point in Π'; that is, Π' is a set of coset representatives for the additive group of complex numbers modulo the subgroup L. We will list zeros and poles in Π', omitting 0 from our list (even if it happens to be a zero or pole of $f(z)$). Each zero or pole will be listed as many times as its multiplicity. However, only "half" will be listed; that is, they will be arranged in pairs, with only one taken from each pair. We now give the details. We describe the method of listing zeros; the method of listing poles is exactly analogous.

First suppose that $a \in \Pi'$, $a \neq 0$, is a zero of $f(z)$ which is *not* half of a lattice point, i.e., $a \neq \omega_1/2$, $\omega_2/2$, or $(\omega_1 + \omega_2)/2$. Let $a^* \in \Pi'$ be the point "symmetric" to a, i.e., $a^* = \omega_1 + \omega_2 - a$ if a is in the interior of Π', while $a^* = \omega_1 - a$ or $a^* = \omega_2 - a$ if a is on one of the two sides (see Fig. I.9). If a is a zero of order m, we claim that the symmetric point a^* is also a zero of order m. This follows from the double periodicity and the evenness of $f(z)$. Namely, we have $f(a^* - z) = f(-a - z)$ by double periodicity, and this is equal to $f(a + z)$ because $f(z)$ is an even function. Thus, if $f(a + z) = a_m z^m +$ higher terms, it follows that $f(a^* + z) = a_m(-z)^m +$ higher terms, i.e., a^* is a zero of order m.

Now suppose that $a \in \Pi'$ is a zero of $f(z)$ which *is* half of a lattice point; for example, suppose that $a = \omega_1/2$. In this case we claim that the order of zero m is even. If $f(a + z) = f(\frac{1}{2}\omega_1 + z) = a_m z^m +$ higher terms, then $f(\frac{1}{2}\omega_1 - z) = f(-\frac{1}{2}\omega_1 + z) = f(\frac{1}{2}\omega_1 + z)$ by double periodicity and evenness. Thus, $a_m z^m +$ higher terms $= a_m(-z)^m +$ higher terms, and so m is even.

We are now ready to list the zeros and poles of $f(z)$. Let $\{a_i\}$ be a list of the zeros of $f(z)$ in Π' which are not half-lattice points, each taken as many times as the multiplicity of zero there, but only one taken from each pair of symmetrical zeros a, a^*; in addition, if one of the three nonzero half-lattice points in Π' is a zero of $f(z)$, include it in the list half as many times as its multiplicity. Let $\{b_j\}$ be a list of the nonzero poles of $f(z)$ in Π', counted in the same way as the zeros (i.e., "only half" of them appear).

Since all of the a_i and b_j are nonzero, the values $\wp(a_i)$ and $\wp(b_j)$ are finite, and it makes sense to define the elliptic function

$$g(z) = \frac{\Pi_i(\wp(z) - \wp(a_i))}{\Pi_j(\wp(z) - \wp(b_j))}.$$

We claim that $g(z)$ has the same zeros and poles as $f(z)$ (counting multiplicity), from which it will follow that $f(z) = c \cdot g(z)$ for some constant c. Since $g(z)$ is a rational function of $\wp(z)$, this will complete the proof.

To prove this claim, we first examine nonzero points in Π'. Since 0 is the only pole in the numerator or denominator of $g(z)$, it follows that the nonzero zeros of $g(z)$ must come from the zeros of $\wp(z) - \wp(a_i)$, while the nonzero poles of $g(z)$ must come from the zeros of $\wp(z) - \wp(b_j)$. But we know (see problems below) that $\wp(z) - u$ (for constant u) has a double zero at $z = u$ if u is a half-lattice point, and otherwise has a pair of simple zeros at u and the symmetric point u^*. These are the only zeros of $\wp(z) - u$ in Π'. By our construction of the a_i and b_j, we see that $g(z)$ and $f(z)$ have the same order of zero or pole everywhere in Π', with the possible exception of the point 0. So it merely remains to show that they have the same order of zero or pole at 0. But this will follow automatically by Proposition 5. Namely, choose α so that no lattice point and no zero or pole of $f(z)$ or $g(z)$ is on the boundary of $\alpha + \Pi$. Then $\alpha + \Pi$ will contain precisely one lattice point l. We know that $f(z)$ and $g(z)$ have the same orders of zeros and poles everywhere in $\alpha + \Pi$ with the possible exception of l. Let m_f denote the order of zero of $f(z)$ at l (m_f is negative if there is a pole), and let m_g denote the analogous order for $g(z)$. Then

$$\begin{aligned} &m_f + (\text{total of orders of zeros of } f) - (\text{total of orders of poles of } f) \\ &\quad = m_g + (\text{total of orders of zeros of } g) - (\text{total of orders of poles of } g). \end{aligned}$$

Since the corresponding terms in parentheses on both sides of the equality are equal, we conclude that $m_f = m_g$. Thus, Proposition 5 tells us that when we know that two elliptic functions have the same order of zero or pole everywhere but possibly at one point in the fundamental parallelogram, then that one point is carried along automatically. This concludes the proof of Proposition 9. □

The proof of Propositions 8 and 9 was constructive, i.e., it gives us a prescription for expressing a given elliptic function in terms of $\wp(z)$ once we know its zeros and poles. Without doing any more work, for example, we can immediately conclude that:

(1) the even elliptic function $\wp'(z)^2$ is a cubic polynomial in $\wp(z)$ (because $\wp'(z)$ has a triple pole at 0 and three simple zeros, hence there are three a_i's and no b_j's);
(2) the even elliptic function $\wp(Nz)$ (for any fixed positive integer N) is a rational function in $\wp(z)$.

Both of these facts will play a fundamental role in what follows. The first tells us that the Weierstrass $\wp$-function satisfies a differential equation of a very special type. This equation will give the connection with elliptic curves. The second fact is the starting point for studying points of finite order on elliptic curves. Both facts will be given a more precise form, and the connection with elliptic curves will be developed, in the sections that follow.

PROBLEMS

1. Prove that the lattice $L = \{m\omega_1 + n\omega_2\}$ and the lattice $L' = \{m\omega_1' + n\omega_2'\}$ are the same if and only if there is a 2×2 matrix A with integer entries and determinant ± 1 such that $\omega' = A\omega$ (where ω denotes the column vector with entries ω_1, ω_2). If the pairs ω_1, ω_2 and ω_1', ω_2' are each listed in clockwise order, show that $\det A = +1$.

2. Let $\mathbb{C}/L$ denote the quotient of the additive group of complex numbers by the subgroup $L = \{m\omega_1 + n\omega_2\}$. Then $\mathbb{C}/L$ is in one-to-one correspondence with the fundamental parallelogram Π with opposite sides glued together.
   (a) Let C be the circle group (the unit circle in the complex plane). Give a continuous group isomorphism from $\mathbb{C}/L$ to the product of C with itself.
   (b) How many points of order N or a divisor of N are there in the group $\mathbb{C}/L$?
   (c) Show that the set of subgroups of prime order p in $\mathbb{C}/L$ is in one-to-one correspondence with the points of $\mathbb{P}^1_{\mathbb{F}_p}$ (where $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$). How many are there?

3. Let $s = 2, 3, 4, \ldots$. Fix a positive integer N, and let $f\colon \mathbb{Z} \times \mathbb{Z} \to \mathbb{C}$ be any function of period N, i.e., $f(m + N, n) = f(m, n)$ and $f(m, n + N) = f(m, n)$. Suppose that $f(0, 0) = 0$. If $s = 2$, further suppose that $\Sigma f(m, n) = 0$, where the sum is over $0 \le m, n < N$. Define a function

$$F_s(\omega_1, \omega_2) = \sum_{m,n \in \mathbb{Z}} \frac{f(m, n)}{(m\omega_1 + n\omega_2)^s}.$$

   (a) Prove that this sum converges absolutely if $s > 2$ and conditionally if $s = 2$ (in the latter case, take the sum over m and n in nondecreasing order of $|m\omega_1 + n\omega_2|$).
   (b) Express $F_s(\omega_1, \omega_2)$ in terms of the values of $\wp(z; \omega_1, \omega_2)$ or a suitable derivative evaluated at values of $z \in \Pi$ for which $Nz \in L$ (see Problem 2(b)).

4. Show that for any fixed u, the elliptic function $\wp(z) - u$ has exactly two zeros (or a single double zero). Use the fact that $\wp'(z)$ is odd to show that the zeros of $\wp'(z)$ are precisely $\omega_1/2$, $\omega_2/2$, and $(\omega_1 + \omega_2)/2$, and that the values $e_1 = \wp(\omega_1/2)$, $e_2 = \wp(\omega_2/2)$, $e_3 = \wp((\omega_1 + \omega_2)/2)$ are the values of u for which $\wp(z) - u$ has a double zero. Why do you know that e_1, e_2, e_3 are distinct? Thus, the Weierstrass $\wp$-function gives a two-to-one map from the torus (the fundamental parallelogram Π with opposite sides glued together) to the Riemann sphere $\mathbb{C} \cup \{\infty\}$ except over the four "branch points" e_1, e_2, e_3, ∞, each of which has a single preimage in $\mathbb{C}/L$.

5. Using the proof of Proposition 9, without doing any computations, what can you say about how the second derivative $\wp''(z)$ can be expressed in terms of $\wp(z)$?

§6. Elliptic curves in Weierstrass form

As remarked at the end of the last section, from the proof of Proposition 9 we can immediately conclude that the square of $\wp'(z)$ is equal to a cubic polynomial in $\wp(z)$. More precisely, we know that $\wp'(z)^2$ has a double zero at $\omega_1/2$, $\omega_2/2$, and $(\omega_1 + \omega_2)/2$ (see Problem 4 of §5). Hence, these three numbers are the a_i's, and we have

$$\begin{aligned}\wp'(z)^2 &= C(\wp(z) - \wp(\omega_1/2))(\wp(z) - \wp(\omega_2/2))(\wp(z) - \wp((\omega_1 + \omega_2)/2))\\ &= C(\wp(z) - e_1)(\wp(z) - e_2)(\wp(z) - e_3),\end{aligned}$$

where C is some constant. It is easy to find C by comparing the coefficients of the lowest power of z in the Laurent expansion at the origin. Recall that $\wp(z) - z^{-2}$ is continuous at the origin, as is $\wp'(z) + 2z^{-3}$. Thus, the leading term on the left is $(-2z^{-3})^2 = 4z^{-6}$, while on the right it is $C(z^{-2})^3 = Cz^{-6}$. We conclude that $C = 4$. That is, $\wp(z)$ satisfies the differential equation

$$\wp'(z)^2 = f(\wp(z)), \qquad \text{where} \quad f(x) = 4(x - e_1)(x - e_2)(x - e_3) \in \mathbb{C}[x]. \tag{6.1}$$

Notice that the cubic polynomial f has distinct roots (see Problem 4 of §5).

We now give another independent derivation of the differential equation for $\wp(z)$ which uses only Proposition 3 from §4. Suppose that we can find a cubic polynomial $f(x) = ax^3 + bx^2 + cx + d$ such that the Laurent expansion at 0 of the elliptic function $f(\wp(z))$ agrees with the Laurent expansion of $\wp'(z)^2$ through the negative powers of z. Then the difference $\wp'(z)^2 - f(\wp(z))$ would be an elliptic function with no pole at zero, or in fact anywhere else (since $\wp(z)$ and $\wp'(z)$ have a pole only at zero). By Proposition 3, this difference is a constant; and if we suitably choose d, the constant term in $f(x)$, we can make this constant zero.

To carry out this plan, we must expand $\wp(z)$ and $\wp'(z)^2$ near the origin. Since both are even functions, only even powers of z will appear.

Let c be the minimum absolute value of nonzero lattice points l. We shall take $r < 1$, and assume that z is in the disc of radius rc about the origin. For each nonzero $l \in L$, we expand the term corresponding to l in the definition (4.1) of $\wp(z)$. We do this by differentiating the geometric series $1/(1 - x) = 1 + x + x^2 + \cdots$ and then substituting z/l for x:

$$\frac{1}{(1 - z/l)^2} = 1 + 2\frac{z}{l} + 3\frac{z^2}{l^2} + 4\frac{z^3}{l^3} + \cdots.$$

If we now subtract 1 from both sides, divide both sides by l^2, and then substitute in (4.1), we obtain

$$\wp(z) = \frac{1}{z^2} + \sum_{\substack{l \in L\\ l \neq 0}} 2\frac{z}{l^3} + 3\frac{z^2}{l^4} + 4\frac{z^3}{l^5} + \cdots + (k - 1)\frac{z^{k-2}}{l^k} + \cdots.$$

We claim that this double series is absolutely convergent for $|z| < rc$, in which case the following reversal of the order of summation will be justified:

$$\wp(z) = \frac{1}{z^2} + 3G_4 z^2 + 5G_6 z^4 + 7G_8 z^6 + \cdots, \tag{6.2}$$

where for $k > 2$ we denote

$$G_k = G_k(L) = G_k(\omega_1, \omega_2) \underset{\text{def}}{=} \sum_{\substack{l \in L \\ l \neq 0}} l^{-k} = \sum_{\substack{m,n \in \mathbb{Z} \\ \text{not both } 0}} \frac{1}{(m\omega_1 + n\omega_2)^k} \tag{6.3}$$

(notice that the G_k are zero for odd k, since the term for l cancels the term for $-l$; as we expect, only even powers of z occur in the expansion (6.2)). To check the claim of absolute convergence of the double series, we write the sum of the absolute values of the terms in the inner sum in the form (recall: $|z| < r|l|$):

$$2|z| \cdot |l|^{-3} \cdot \left(1 + \frac{3}{2}r + \frac{4}{2}r^2 + \frac{5}{2}r^3 + \cdots\right) < \frac{2|z|}{(1-r)^2}\frac{1}{|l|^3},$$

and then use Lemma 2 from the proof of Proposition 6.

We now use (6.2) to compute the first few terms in the expansions of $\wp(z)$, $\wp(z)^2$, $\wp(z)^3$, $\wp'(z)$, and $\wp'(z)^2$, as follows:

$$\wp'(z) = -\frac{2}{z^3} + 6G_4 z + 20G_6 z^3 + 42G_8 z^5 + \cdots; \tag{6.4}$$

$$\wp'(z)^2 = \frac{4}{z^6} - 24G_4\frac{1}{z^2} - 80G_6 + (36G_4^2 - 168G_8)z^2 + \cdots; \tag{6.5}$$

$$\wp(z)^2 = \frac{1}{z^4} + 6G_4 + 10G_6 z^2 + \cdots; \tag{6.6}$$

$$\wp(z)^3 = \frac{1}{z^6} + 9G_4\frac{1}{z^2} + 15G_6 + (21G_8 + 27G_4^2)z^2 + \cdots. \tag{6.7}$$

Recall that we are interested in finding coefficients a, b, c, d of a cubic $f(x) = ax^3 + bx^2 + cx + d$ such that

$$\wp'(z)^2 = a\wp(z)^3 + b\wp(z)^2 + c\wp(z) + d,$$

and we saw that it suffices to show that both sides agree in their expansion through the constant term. If we multiply equation (6.7) by a, equation (6.6) by b, equation (6.2) by c, and then add them all to the constant d, and finally equate the coefficients of z^{-6}, z^{-4}, z^{-2} and the constant term to the corresponding coefficients in (6.5), we obtain successively:

$$a = 4; \qquad b = 0; \qquad -24G_4 = 4(9G_4) + c; \qquad -80G_6 = 4(15G_6) + d.$$

Thus, $c = -60G_4$, $d = -140G_6$. It is traditional to denote

$$g_2 = g_2(L) \underset{\text{def}}{=} 60G_4 = 60 \sum_{\substack{l \in L \\ l \neq 0}} l^{-4};$$
$$g_3 = g_3(L) \underset{\text{def}}{=} 140G_6 = 140 \sum_{\substack{l \in L \\ l \neq 0}} l^{-6}. \tag{6.8}$$

We have thereby derived a second form for the differential equation (6.1):

$$\wp'(z)^2 = f(\wp(z)), \qquad \text{where} \quad f(x) = 4x^3 - g_2 x - g_3 \in \mathbb{C}[x]. \tag{6.9}$$

Notice that if we were to continue comparing coefficients of higher powers of z in the expansion of both sides of (6.9), we would obtain relations between the various G_k (see Problems 4–5 below).

The differential equation (6.9) has an elegant and basic geometric interpretation. Suppose that we take the function from the torus $\mathbb{C}/L$ (i.e., the fundamental parallelogram Π with opposite sides glued) to $\mathbb{P}^2_{\mathbb{C}}$ defined by

$$z \mapsto (\wp(z), \wp'(z), 1) \quad \text{for} \quad z \neq 0;$$
$$0 \mapsto (0, 1, 0). \tag{6.10}$$

The image of any nonzero point z of $\mathbb{C}/L$ is a point in the xy-plane (with complex coordinates) whose x- and y-coordinates satisfy the relationship $y^2 = f(x)$ because of (6.9). Here $f(x) \in \mathbb{C}[x]$ is a cubic polynomial with distinct roots. Thus, every point z in $\mathbb{C}/L$ maps to a point on the elliptic curve $y^2 = f(x)$ in $\mathbb{P}^2_{\mathbb{C}}$. It is not hard to see that this map is a one-to-one correspondence between $\mathbb{C}/L$ and the elliptic curve (including its point at infinity). Namely, every x-value except for the roots of $f(x)$ (and infinity) has precisely two z's such that $\wp(z) = x$ (see Problem 4 of §5). The y-coordinates $y = \wp'(z)$ coming from these two z's are the two square roots of $f(x) = f(\wp(z))$. If, however, x happens to be a root of $f(x)$, then there is only one z value such that $\wp(z) = x$, and the corresponding y-coordinate is $y = \wp'(z) = 0$, so that again we are getting the solutions to $y^2 = f(x)$ for our given x.

Moreover, the map from $\mathbb{C}/L$ to our elliptic curve in $\mathbb{P}^2_{\mathbb{C}}$ is analytic, meaning that near any point of $\mathbb{C}/L$ it can be given by a triple of analytic functions. Near non-lattice points of $\mathbb{C}$ the map is given by $z \mapsto (\wp(z), \wp'(z), 1)$; and near lattice points the map is given by $z \mapsto (\wp(z)/\wp'(z), 1, 1/\wp'(z))$, which is a triple of analytic functions near L.

We have proved the following proposition.

Proposition 10. *The map* (6.10) *is an analytic one-to-one correspondence between* $\mathbb{C}/L$ *and the elliptic curve* $y^2 = 4x^3 - g_2(L)x - g_3(L)$ *in* $\mathbb{P}^2_{\mathbb{C}}$.

One might be interested in how the inverse map from the elliptic curve to $\mathbb{C}/L$ can be constructed. This can be done by taking path integrals of $dx/y = (4x^3 - g_2 x - g_3)^{-1/2} dx$ from a fixed starting point to a variable endpoint. The resulting integral depends on the path, but only changes by

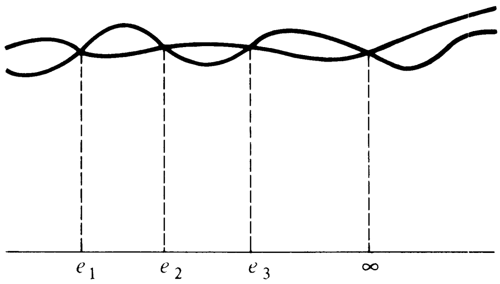


Figure I.10

a "period", i.e., a lattice element, if we change the path. We hence obtain a well-defined map to $\mathbb{C}/L$. See the exercises below for more details.

We conclude this section with a few words about an algebraic picture that is closely connected with the geometric setting of our elliptic curve. Recall from Proposition 8 that any elliptic function (meromorphic function on the torus $\mathbb{C}/L$) is a rational expression in $\wp(z)$ and $\wp'(z)$. Under our one-to-one correspondence in Proposition 10, such a function is carried over to a rational expression in x and y on the elliptic curve in the xy-plane (actually, in $\mathbb{P}^2_{\mathbb{C}}$). Thus, the field $\mathbb{C}(x, y)$ of rational functions on the xy-plane, when we restrict its elements to the elliptic curve $y^2 = f(x)$, and then "pull back" to the torus $\mathbb{C}/L$ by substituting $x = \wp(z)$, $y = \wp'(z)$, give us precisely the elliptic functions $\mathscr{E}_L$. Since the restriction of y^2 is the same as the restriction of $f(x)$, the field of functions obtained by restricting the rational functions in $\mathbb{C}(x, y)$ to the elliptic curve is the following quadratic extension of $\mathbb{C}(x)$: $\mathbb{C}(x)[y]/(y^2 - (4x^3 - g_2x - g_3))$. Algebraically speaking, we form the quotient ring of $\mathbb{C}(x)[y]$ by the principal ideal corresponding to the equation $y^2 = f(x)$.

Geometrically, projection onto the x-coordinate gives us Fig. I.10. Two points on the elliptic curve map to one point on the projective line, except at four points (the point at infinity and the three points where $y = 0$), where the two "branches" are "pinched" together.

In algebraic geometry, one lets the field $F = \mathbb{C}(x)$ correspond to the complex line $\mathbb{P}^1_{\mathbb{C}}$, and the field $K = \mathbb{C}(x, y)/(y^2 - (4x^3 - g_2x - g_3))$ correspond to the elliptic curve in $\mathbb{P}^2_{\mathbb{C}}$. The rings $A = \mathbb{C}[x]$ and $B = \mathbb{C}[x, y]/(y^2 - f(x))$ are the "rings of integers" in these fields. The maximal ideals in A are of the form $(x - a)A$; they are in one-to-one correspondence with $a \in \mathbb{C}$. A maximal ideal in B is of the form $(x - a)B + (y - b)B$ (where b is a square root of $f(a)$), and it corresponds to the point (a, b) on the elliptic curve.

$$\begin{array}{ccccl} K & \supset & B & \supset & (x-a)B + (y-b)B \qquad (b = \sqrt{f(a)}) \\ | & & | & & \\ | & & | & & (x-a)B + (y+b)B \\ | & & | & & \qquad | \\ F & \supset & A & \supset & (x-a)A \end{array}$$

The maximal ideal $(x - a)A$, when "lifted up" to the ring B, is no longer prime. That is, the ideal $(x - a)B$ factors into the product of the two ideals:

$$(x - a)B = ((x - a)B + (y - b)B)((x - a)B + (y + b)B).$$

The maximal ideal corresponding to the point a on the x-line splits into two maximal ideals corresponding to two points on the elliptic curve. If it so happens that $b = 0$, i.e., a is a root of $f(x)$, then both of the ideals are the same, i.e., $(x - a)B$ is the square of the ideal $((x - a)B + yB)$. In that case we say that the ideal $(x - a)A$ "ramifies" in B. This happens at values a of the x-coordinate which come from only one point $(a, 0)$ on the elliptic curve. Thus, the above algebraic diagram of fields, rings and ideals is an exact mirror of the preceding geometric diagram.

We shall not go further than these *ad hoc* comments, since we shall not be using algebra geometric techniques in which follows. For a systematic introduction to algebraic geometry, see the textbooks by Shafarevich, Mumford, or Hartshorne.

PROBLEMS

1. (a) Let $L = \mathbb{Z}[i]$ be the lattice of Gaussian integers. Show that $g_3(L) = 0$ but that $g_2(L)$ is a nonzero real number.
   (b) Let $L = \mathbb{Z}[\omega]$, where $\omega = \frac{1}{2}(-1 + i\sqrt{3})$, be the lattice of integers in the quadratic imaginary field $\mathbb{Q}(\sqrt{-3})$. Show that $g_2(L) = 0$ but that $g_3(L)$ is a nonzero real number.
   (c) For any nonzero complex number c, let cL denote the lattice obtained by multiplying all lattice elements by c. Show that $g_2(cL) = c^{-4}g_2(L)$, and $g_3(cL) = c^{-6}g_3(L)$.
   (d) Prove that any elliptic curve $y^2 = 4x^3 - g_2x - g_3$ with either g_2 or g_3 equal to zero, is of the form $y^2 = 4x^3 - g_2(L)x - g_3(L)$ for some lattice L. It can be shown that any elliptic curve is of that form for some lattice L. See, for example, [Whittaker & Watson 1958, §21.73]; also, we shall prove this much later as a corollary in our treatment of modular forms.

2. Recall that the discriminant of a polynomial $f(x) = a_0x^n + a_1x^{n-1} + \cdots + a_n = a_0(x - e_1))x - e_2)\cdots(x - e_n)$ is $a_0^{n-1}\Pi_{i<j}(e_i - e_j)^2$. It is nonzero if and only if the roots are distinct. Since it is a symmetric homogeneous polynomial of degree $n(n - 1)$ in the $e_i's$, it can be written as a polynomial in the elementary symmetric polynomials in the $e_i's$, which are $(-1)^i a_i/a_0$. Moreover, each monomial term $\Pi_i (a_i/a_0)^{m_i}$ has total "weight" $m_1 + 2m_2 + \cdots + nm_n$ equal to $n(n - 1)$. Applying this to $f(x) = 4x^3 - g_2x - g_3$, we see that the discriminant is equal to a polynomial in g_2, g_3 of weight six, i.e., it must be of the form $\alpha g_2^3 + \beta g_3^2$. Find α and β by computing $4^2(e_1 - e_2)^2(e_1 - e_3)^2(e_2 - e_3)^2$ directly in the case $g_2 = 4$, $g_3 = 0$ and the case $g_2 = 0$, $g_3 = 4$.

3. Since the even elliptic function $\wp''(z)$ has a quadruple pole at zero and no other pole, you know in advance that it is equal to a quadratic polynomial in $\wp(z)$. Find this polynomial in two ways: (a) comparing coefficients of powers of z; (b) differentiating $\wp'^2 = 4\wp^3 - g_2\wp - g_3$. Check that your answers agree.

4. Use either the equation for $\wp'^2$ or the equation for $\wp''$ to prove that $G_8 = \frac{3}{7}G_4^2$.

5. Prove by induction that all G_k's can be expressed as polynomials in G_4 and G_6 with rational coefficients, i.e., $G_k \in \mathbb{Q}[G_4, G_6]$. We shall later derive this fact again when we study modular forms (of which the G_k turn out to be examples).

6. Let $\omega_1 = it$ be purely imaginary, and let $\omega_2 = \pi$. Show that as t approaches infinity, $G_k(it, \pi)$ approaches $2\pi^{-k}\zeta(k)$, where $\zeta(s)$ is the Riemann zeta-function. Suppose we know that $\zeta(2) = \pi^2/6$, $\zeta(4) = \pi^4/90$, $\zeta(6) = \pi^6/945$. Use Problem 4 to find $\zeta(8)$. Use Problem 5 to show that $\pi^{-k}\zeta(k) \in \mathbb{Q}$ for all positive even integers k.

7. Find the limit of g_2 and g_3 for the lattice $L = \{mit + n\pi\}$ as $t \to \infty$.

8. Show that $v = \csc^2 z$ satisfies the differential equation $v'^2 = 4v^2(v - 1)$, and that the function
$$v = \csc^2 z - \tfrac{1}{3}$$
satisfies the differential equation $v'^2 = 4v^3 - \frac{4}{3}v - \frac{8}{27}$. What is the discriminant of the polynomial on the right? Now start with the infinite product formula for $\sin(\pi z)$, replace z by z/π, and take the logarithmic derivative and then the derivative once again to obtain an infinite sum for $\csc^2 z$. Then prove that
$$\lim_{t\to\infty} \wp(z; it, \pi) = \csc^2 z - \tfrac{1}{3}.$$

9. The purpose of this problem is to review the function $z = \log v$ for v complex, in the process providing a "dry run" for the problems that follow.
   (a) For v in a simply connected region of the complex plane that does not include the origin, define a function z of v by:
$$z = \int_1^v \frac{dt}{t},$$
   where the path from 1 to v is chosen arbitrarily, except that the same choice is made for all points in the region. (In other words, fix any path from 1 to v_0, and then to go to other v's use a path from v_0 to v that stays in the region.) Call this function $z = \log v$. Show that if a different path is chosen, the function changes by a constant value in the "lattice" $L = \{2\pi i m\}$; and that any lattice element can be added to the function by a suitable change of path. (L is actually only a lattice in the imaginary axis $\mathbb{R}i$, not a lattice in $\mathbb{C}$.)
   (b) Express dz/dv and dv/dz in terms of v.
   (c) If the function $v = e^z$ is defined by the usual series, use part (b) to show that e^z is the inverse function of $z = \log v$.
   (d) Show that the map e^z gives a one-to-one correspondence between $\mathbb{C}/L$ and $\mathbb{C} - \{0\}$. Under this one-to-one correspondence, the additive group law in $\mathbb{C}/L$ becomes what group law in $\mathbb{C} - \{0\}$?

10. Let L be a fixed lattice, set $g_2 = g_2(L)$, $g_3 = g_3(L)$, $\wp(z) = \wp(z; L)$. Let $u = f(z)$ be a non-constant function on a connected open region $R \subset \mathbb{C}$ which satisfies the differential equation $u'^2 = 4u^3 - g_2u - g_3$. Prove that $u = \wp(z + \alpha)$ for some constant α.

11. Let $L = \{m\omega_1 + n\omega_2\}$ be a fixed lattice, and set $g_2 = g_2(L)$, $g_3 = g_3(L)$, $\wp(z) = \wp(z; L)$. Let R_1 be an unbounded simply connected open region in the complex plane which does not contain the roots e_1, e_2, e_3 of the cubic $4x^3 - g_2x - g_3$.

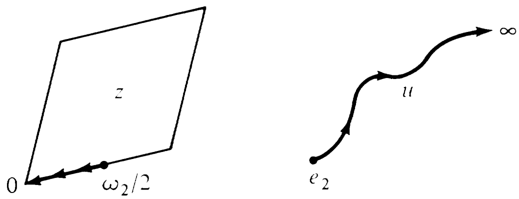


Figure I.11

For $u \in R_1$, define a function $z = g(u)$ by

$$z = g(u) = \int_u^\infty \frac{dt}{\sqrt{4t^3 - g_2 t - g_3}},$$

where a fixed branch of the square root is chosen as t varies in R_1. Note that the integral converges and is independent of the path in R_1 from u to ∞, since R_1 is simply connected. The function $z = g(u)$ can be analytically continued by letting R_2 be a simply connected region in $\mathbb{C} - \{e_1, e_2, e_3\}$ which overlaps with R_1. If $u \in R_2$, then choose $u_1 \in R_1 \cap R_2$, and set $z = g(u) = g(u_1) + \int_u^{u_1} (4t^3 - g_2 t - g_3)^{-1/2} dt$. This definition clearly does not depend on our choice of $u_1 \in R_1 \cap R_2$ or our path from u to u_1 in R_2. Continuing in this way, we obtain an analytic function which is multivalued, because our sequence of regions $R_1, R_2, R_3, \ldots$ can wind around e_1, e_2, or e_3.

(a) Express $(dz/du)^2$ and $(du/dz)^2$ in terms of u.

(b) Show that $u = \wp(z)$. In particular, when we wind around e_1, e_2, or e_3 the value of z can only change by something in L. Thus, $z = g(u)$ is well defined as an element in $\mathbb{C}/L$ for $u \in \mathbb{C} - \{e_1, e_2, e_3\}$. The function $z = g(u)$ then extends by continuity to e_1, e_2, e_3.

(c) Let C_1 be the path in the complex u-plane from e_2 to ∞ that is traced by $u = \wp(z)$ as z goes from $\omega_2/2$ to 0 along the side of Π (see Fig. I.11). Show that $\int_{C_1} (4t^3 - g_2 t - g_3)^{-1/2} dt = -\omega_2/2$ for a suitable branch of the square root.

(d) Let C_2 be the path that goes from ∞ to e_2 along C_1, winds once around e_2, and then returns along C_1 to ∞. Take the same branch of the square root as in part (c), and show that $\int_{C_2} (4t^3 - g_2 t - g_3)^{-1/2} dt = \omega_2$.

(e) Describe how the function $z = g(u)$ can be made to give all preimages of u under $u = \wp(z)$.

12. (a) Prove that all of the roots e_1, e_2, e_3 of $4x^3 - g_2 x - g_3$ are real if and only if g_2 and g_3 are real and $\Delta = g_2^3 - 27g_3^2 > 0$.

(b) Suppose that the conditions in part (a) are met, and we order the e_i so that $e_2 > e_3 > e_1$. Show that we can choose the periods of L to be given by

$$\frac{1}{2}\omega_1 = i \int_{-\infty}^{e_1} \frac{dt}{\sqrt{g_3 + g_2 t - 4t^3}} \quad \text{and} \quad \frac{1}{2}\omega_2 = \int_{e_2}^{\infty} \frac{dt}{\sqrt{4t^3 - g_2 t - g_3}},$$

where we take the positive branch of the square root, and integrate along the real axis.

(c) With these assumptions about the location of the e_i on the real axis, describe how to change the path of integration and the branch of the square root in

Problem 11 so as to get the other values of z for which $u = \wp(z)$, namely $\pm z + m\omega_1 + n\omega_2$.

13. Suppose that $g_2 = 4n^2$, $g_3 = 0$. Take e_1, e_2, e_3 so that $e_2 > e_3 > e_1$. What are e_1, e_2, e_3 in this case? Show that $\omega_1 = i\omega_2$, i.e., the lattice L is the Gaussian integer lattice expanded by a factor of ω_2. Show that as z travels along the straight line from $\omega_1/2$ to $\omega_1/2 + \omega_2$ the point $(x, y) = (\wp(z), \wp'(z))$ moves around the real points of the elliptic curve $y^2 = 4(x^3 - n^2x)$ between $-n$ and 0; and as z travels along the straight line from 0 to ω_2 the point $(x, y) = (\wp(z), \wp'(z))$ travels through all the real points of this elliptic curve which are to the right of $(n, 0)$. Think of the "open" appearance of the latter path to be an optical illusion: the two ends are really "tied together" at the point at infinity (0, 1, 0).

14. (a) Show that $\int_0^1 \frac{t^n dt}{\sqrt{t(1-t)}} = \frac{\pi}{n!} \cdot \frac{1}{2} \cdot \frac{3}{2} \cdot \frac{5}{2} \cdots \left(n - \frac{1}{2}\right)$ for $n = 0, 1, 2, \ldots$.

(b) Under the conditions of Problem 12, with $e_2 > e_3 > e_1$, set $\lambda = \frac{e_3 - e_1}{e_2 - e_1} \in (0, 1)$. Derive the formula:

$$\omega_2 = \frac{1}{\sqrt{e_2 - e_1}} \int_0^1 \frac{dt}{\sqrt{t(1-t)(1-\lambda t)}}.$$

(c) Derive the formula $\omega_2 = \pi(e_2 - e_1)^{-1/2}F(\lambda)$, where

$$F(\lambda) = \sum_{n=0}^{\infty} \left[\frac{1}{2} \cdot \frac{3}{2} \cdot \frac{5}{2} \cdots \left(n - \frac{1}{2}\right)\right]^2 \frac{\lambda^n}{n!^2}.$$

The function $F(\lambda)$ is called a "hypergeometric series".

(d) Show that the hypergeometric series in part (c) satisfies the differential equation: $\lambda(1 - \lambda)F''(\lambda) + (1 - 2\lambda)F'(\lambda) - \frac{1}{4}F(\lambda) = 0$.

§7. The addition law

In the last section we showed how the Weierstrass $\wp$-function gives a correspondence between the points of $\mathbb{C}/L$ and the points on the elliptic curve $y^2 = f(x) = 4x^3 - g_2(L)x - g_3(L)$ in $\mathbb{P}^2_{\mathbb{C}}$. We have an obvious addition law for points in $\mathbb{C}/L$, obtained from ordinary addition of complex numbers by dividing by the additive subgroup L, i.e., ordinary addition "modulo L". This is the two-dimensional analog of "addition modulo one" in the group $\mathbb{R}/\mathbb{Z}$.

We can use the correspondence between $\mathbb{C}/L$ and the elliptic curve to carry over the addition law to the points on the elliptic curve. That is, to add two points $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$, by definition what we do is go back to the z-plane, find z_1 and z_2 such that $P_1 = (\wp(z_1), \wp'(z_1))$ and $P_2 = (\wp(z_2), \wp'(z_2))$, and then set $P_1 + P_2 = (\wp(z_1 + z_2), \wp'(z_1 + z_2))$. This is just a case of the general principle: whenever we have a one-to-one correspondence between elements of a commutative group and elements of some

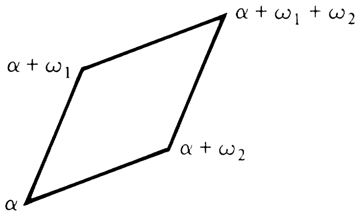


Figure I.12

other set, we can use this correspondence to define a commutative group law on that other set.

But the remarkable thing about the addition law we obtain in this way is that (1) there is a simple geometric interpretation of "adding" the points on the elliptic curve, and (2) the coordinates of $P_1 + P_2$ can be expressed directly in terms of x_1, x_2, y_1, y_2 by rather simple rational functions. The purpose of this section is to show how this is done.

We first prove a general lemma about elliptic functions.

Lemma. *Let $f(z) \in \mathscr{E}_L$. Let $\Pi = \{a\omega_1 + b\omega_2 | 0 \le a, b \le 1\}$ be a fundamental parallelogram for the lattice L, and choose α so that $f(z)$ has no zeros or poles on the boundary of $\alpha + \Pi$. Let $\{a_i\}$ be the zeros of $f(z)$ in $\alpha + \Pi$, each repeated as many times as its multiplicity, and let $\{b_j\}$ be the poles, each occurring as many times as its multiplicity. Then $\Sigma a_i - \Sigma b_j \in L$.*

PROOF. Recall that the function $f'(z)/f(z)$ has poles at the zeros and poles of $f(z)$, and its expansion near a zero a of order m is $m/(z - a) + \cdots$ (and near a pole b of order $-m$ the expansion is $-m/(z - b) + \cdots$). Then the function $zf'(z)/f(z)$ has the same poles, but, writing $z = a + (z - a)$, we see that the expansion starts out $am/(z - a)$. We conclude that $\Sigma a_i - \Sigma b_j$ is the sum of the residues of $zf'(z)/f(z)$ inside $\alpha + \Pi$. Let C be the boundary of $\alpha + \Pi$. By the residue theorem,

$$\sum a_i - \sum b_j = \frac{1}{2\pi i} \int_C \frac{zf'(z)}{f(z)} dz.$$

We first take the integral over the pair of opposite sides from α to $\alpha + \omega_2$ and from $\alpha + \omega_1$ to $\alpha + \omega_1 + \omega_2$ (see Fig. I.12). This part is equal to

$$\frac{1}{2\pi i}\left(\int_\alpha^{\alpha+\omega_2} z\frac{f'(z)}{f(z)}dz - \int_{\alpha+\omega_1}^{\alpha+\omega_1+\omega_2} z\frac{f'(z)}{f(z)}dz\right)$$
$$= \frac{1}{2\pi i}\left(\int_\alpha^{\alpha+\omega_2} z\frac{f'(z)}{f(z)}dz - \int_\alpha^{\alpha+\omega_2} (z + \omega_1)\frac{f'(z)}{f(z)}dz\right)$$
$$= -\omega_1 \frac{1}{2\pi i}\int_\alpha^{\alpha+\omega_2} \frac{f'(z)}{f(z)}dz.$$

Now make the change of variables $u = f(z)$, so that $f'(z)dz/f(z) = du/u$. Let C_1 be the closed path from $f(\alpha)$ to $f(\alpha + \omega_2) = f(\alpha)$ traced by $u = f(z)$ as z goes from α to $\alpha + \omega_2$. Then

$$\frac{1}{2\pi i}\int_{\alpha}^{\alpha+\omega_2} \frac{f'(z)}{f(z)}dz = \frac{1}{2\pi i}\int_{C_1} \frac{du}{u},$$

and this is some integer n, namely the number of times the closed path C_1 winds around the origin (counterclockwise). Thus, we obtain $-\omega_1 n$ for this part of our original integral. In the same way, we find that the integral over the remaining two sides of C is equal to $-\omega_2 m$ for some integer m. Thus, $\Sigma\, a_i - \Sigma\, b_j = -n\omega_1 - m\omega_2 \in L$, as desired. This proves the lemma. □

We are now ready to derive the geometrical procedure for adding two points on the elliptic curve $y^2 = f(x) = 4x^3 - g_2(L)x - g_3(L)$. For z in $\mathbb{C}/L$, let P_z be the corresponding point $P_z = (\wp(z), \wp'(z), 1)$, $P_0 = (0, 1, 0)$ on the elliptic curve. Suppose we want to add $P_{z_1} = (x_1, y_1)$ to $P_{z_2} = (x_2, y_2)$ to obtain the sum $P_{z_1+z_2} = (x_3, y_3)$. We would like to know how to go from the two points to their sum directly, without tracing the points back to the z-plane.

We first treat some special cases. The additive identity is, of course, the image of $z = 0$. Let 0 denote the point at infinity $(0, 1, 0)$, i.e., the additive identity of our group of points. The addition is trivial if one of the points is 0, i.e., if z_1 or z_2 is zero. Next, suppose that P_{z_1} and P_{z_2} have the same x-coordinate but are not the same point. This means that $x_2 = x_1, y_2 = -y_1$. In this case $z_2 = -z_1$, because only "symmetric" values of z (values which are the negatives of each other modulo the lattice L) can have the same $\wp$-value. In this case, $P_{z_1} + P_{z_2} = P_0 = 0$, i.e., the two points are additive inverses of one another. Speaking geometrically, we say that two points of the curve which are on the same vertical line have sum 0. We further note that in the special situation of a point $P_{z_1} = P_{z_2}$ on the x-axis, we have $y_2 = -y_1 = 0$, and it is easy to check that we still have $P_{z_1} + P_{z_2} = 2P_{z_1} = 0$. We have proved:

Proposition 11. *The additive inverse of* (x, y) *is* $(x, -y)$.

Given two points $P_1 = P_{z_1} = (x_1, y_1)$ and $P_2 = P_{z_2} = (x_2, y_2)$ on the elliptic curve $y^2 = 4x^3 - g_2 x - g_3$ (neither the point at infinity 0), there is a line $l = \overline{P_1 P_2}$ joining them. If $P_1 = P_2$, we take l to be the tangent line to the elliptic curve at P_1. If l is a vertical line, then we saw that $P_1 + P_2 = 0$. Suppose that l is not a vertical line, and we want to find $P_1 + P_2 = P_3 = (x_3, y_3)$. Our basic claim is that $-P_3 = (x_3, -y_3)$ is the third point of intersection of the elliptic curve with l.

Write the equation of $l = \overline{P_1 P_2}$ in the form $y = mx + b$. A point (x, y) on l is on the elliptic curve if and only if $(mx + b)^2 = f(x) = 4x^3 - g_2 x - g_3$, that is, if and only if x is a root of the cubic $f(x) - (mx + b)^2$. This cubic

has three roots, each of which gives a point of intersection. If x is a double root or triple root, then l intersects the curve with multiplicity two or three at the point (x, y) (see Problem 6 of §I.3). In any case, the total number of points of intersection (counting multiplicity) is three.

Notice that vertical lines also intersect the curve in three points, including the point at infinity 0; and the line at infinity has a triple intersection at 0 (see Problem 7 of §I.3). Thus, any line in $\mathbb{P}^2_{\mathbb{C}}$ intersects the curve in three points. This is a special case of

Bezout's Theorem. *Let $\tilde{F}(x, y, z)$ and $\tilde{G}(x, y, z)$ be homogeneous polynomials of degree m and n, respectively, over an algebraically closed field K. Suppose that $\tilde{F}$ and $\tilde{G}$ have no common polynomial factor. Then the curves in $\mathbb{P}^2_K$ defined by $\tilde{F}$ and $\tilde{G}$ have mn points of intersection, counting multiplicities.*

For a more detailed discussion of multiplicity of intersection and a proof of Bezout's theorem, see, for example, Walker's book an algebraic curves [Walker 1978].

In our case $\tilde{F}(x, y, z) = y^2z - 4x^3 + g_2xz^2 + g_3z^3$ and $\tilde{G}(x, y, z) = y - mx - bz$.

Proposition 12. *If $P_1 + P_2 = P_3$, then $-P_3$ is the third point of intersection of $l = \overline{P_1P_2}$ with the elliptic curve. If $P_1 = P_2$, then by $\overline{P_1P_2}$ we mean the tangent line at P_1.*

PROOF. We have already treated the case when P_1 or P_2 is the point at infinity 0, and when $P_2 = -P_1$. So suppose that $l = \overline{P_1P_2}$ has the form $y = mx + b$. Let $P_1 = P_{z_1}$, $P_2 = P_{z_2}$. To say that a point $P_z = (\wp(z), \wp'(z))$ is on l means that $\wp'(z) = m\wp(z) + b$. The elliptic function $\wp'(z) - m\wp(z) - b$ has three poles and hence three zeros in $\mathbb{C}/L$. Both z_1 and z_2 are zeros. According to the lemma proved above, the sum of the three zeros and three poles is equal to zero modulo the lattice L. But the three poles are all at zero (where $\wp'(z)$ has a triple pole); thus, the third zero is $-(z_1 + z_2)$ modulo the lattice. Hence, the third point of intersection of l with the curve is $P_{-(z_1+z_2)} = -P_{z_3}$, as claimed.

The argument in the last paragraph is rigorous only if the three points of intersection of l with the elliptic curve are distinct, in which case a zero of $\wp'(z) - m\wp(z) - b$ corresponds exactly to a point of intersection P_z. Otherwise, we must show that a double or triple zero of the elliptic function always corresponds to a double or triple intersection, respectively, of l with the curve. That is, we must show that the two meanings of the term "multiplicity" agree: multiplicity of zero of the elliptic function of the variable z, and multiplicity of intersection in the xy-plane.

Let z_1, z_2, $-z_3$ be the three zeros of $\wp'(z) - m\wp(z) - b$, listed as many times as their multiplicity. Note that none of these three points is the negative of another one, since l is not a vertical line. Since $-z_1$, $-z_2$, z_3 are the three

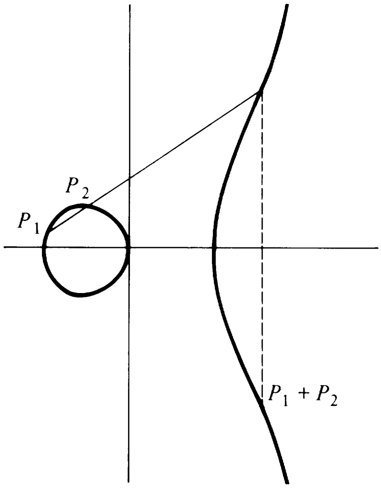


Figure I.13

zeros of $\wp'(z) + m\wp(z) + b$, it follows that $\pm z_1$, $\pm z_2$, $\pm z_3$ are the six zeros of $\wp'(z)^2 - (m\wp(z) + b)^2 = f(\wp(z)) - (m\wp(z) + b)^2 = 4(\wp(z) - x_1)(\wp(z) - x_2)(\wp(z) - x_3)$, where x_1, x_2, x_3 are the roots of $f(x) - (mx + b)^2$. If, say, $\wp(z_1) = x_1$, then the multiplicity of x_1 depends upon the number of $\pm z_2$, $\pm z_3$ which equal $\pm z_1$. But this is precisely the number of z_2, $-z_3$ which equal z_1. Hence "multiplicity" has the same meaning in both cases.

This concludes the proof of Proposition 12. □

Proposition 12 gives us Fig. I.13, which illustrates the group of real points on the elliptic curve $y^2 = x^3 - x$. To add two points P_1 and P_2, we draw the line joining them, find the third point of intersection of that line with the curve, and then take the symmetric point on the other side of the x-axis.

It would have been possible to define the group law in this geometrical manner in the first place, and prove directly that the axioms of an abelian group are satisfied. The hardest part would have been the associative law, which would have necessitated a deeper investigation of intersections of curves. In turns out that there is some flexibility in defining the group law. For example, any one of the eight points of inflection besides the point at infinity could equally well have been chosen as the identity. For details of this alternate approach, see [Walker 1978].

One disadvantage of our approach using $\wp(z)$ is that *a priori* it only applies to elliptic curves of the form $y^2 = 4x^3 - g_2(L)x - g_3(L)$ or curves that can be transformed to that form by a linear change of variables. (Note that the geometrical description of the group law will still give an abelian group law after a linear change of variables.) In actual fact, as was mentioned earlier and will be proved later, any elliptic curve over the complex numbers can be transformed to the Weierstrass form for some lattice L. We already know

that our favorite example $y^2 = x^3 - n^2x$ corresponds to a multiple of the Gaussian integer lattice. In the exercises for this section and the next, we shall allow ourselves to use the fact that the group law works for any elliptic curve.

It is not hard to translate this geometrical procedure into formulas expressing the coordinates (x_3, y_3) of the sum of $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$ in terms of x_1, x_2, y_1, y_2 and the coefficients of the equation of the elliptic curve. Although, strictly speaking, our derivation was for elliptic curves in the form $y^2 = f(x) = 4x^3 - g_2(L)x - g_3(L)$ for some lattice L, the procedure gives an abelian group law for any elliptic curve $y^2 = f(x)$, as remarked above. So let us take $f(x) = ax^3 + bx^2 + cx + d \in \mathbb{C}[x]$ to be any cubic with distinct roots.

In what follows, we shall assume that neither P_1 nor P_2 is the point at infinity 0, and that $P_1 \neq -P_2$. Then the line through P_1 and P_2 (the tangent line at P_1 if $P_1 = P_2$) can be written in the form $y = mx + \beta$, where $m = (y_2 - y_1)/(x_2 - x_1)$ if $P_1 \neq P_2$ and $m = dy/dx|_{(x_1, y_1)}$ if $P_1 = P_2$. In the latter case we can express m in terms of x_1 and y_1 by implicitly differentiating $y^2 = f(x)$; we find that $m = f'(x_1)/2y_1$. In both cases the y-intercept is $\beta = y_1 - mx_1$.

Then x_3, the x-coordinate of the sum, is the third root of the cubic $f(x) - (mx + \beta)^2$, two of whose roots are x_1, x_2. Since the sum of the three roots is equal to minus the coefficient of x^2 divided by the leading coefficient, we have: $x_1 + x_2 + x_3 = -(b - m^2)/a$, and hence:

$$x_3 = -x_1 - x_2 - \frac{b}{a} + \frac{1}{a}\left(\frac{y_2 - y_1}{x_2 - x_1}\right)^2, \qquad \text{if} \quad P_1 \neq P_2; \tag{7.1}$$

$$x_3 = -2x_1 - \frac{b}{a} + \frac{1}{a}\left(\frac{f'(x_1)}{2y_1}\right)^2, \qquad \text{if} \quad P_1 = P_2. \tag{7.2}$$

The y-coordinate y_3 is the negative of the value $y = mx_3 + \beta$, i.e.,

$$y_3 = -y_1 + m(x_1 - x_3), \tag{7.3}$$

where x_3 is given by (7.1) and (7.2), and

$$\begin{aligned} m &= (y_2 - y_1)/(x_2 - x_1) \quad &&\text{if} \quad P_1 \neq P_2; \\ m &= f'(x_1)/2y_1 &&\text{if} \quad P_1 = P_2. \end{aligned} \tag{7.4}$$

If our elliptic curve is in Weierstrass form $y^2 = 4x^3 - g_2x - g_3$, then we have $a = 4$, $b = 0$, and $f'(x_1) = 12x_1^2 - g_2$ in the addition formulas (7.1)–(7.4).

In principle, we could have simply defined the group law by means of these formulas, and then verified algebraically that the axioms of a commutative group are satisfied. The hardest axiom to verify would be associativity. Tedious as this procedure would be, it would have one key advantage over

either the complex-analytic procedure (using $\wp(z)$) or the geometrical procedure. Namely, we would never have to use the fact that our field K over which the elliptic curve is defined is the complex numbers, or even that it has characteristic zero. That is, we would find that our formulas, which make sense over any field K of characteristic not equal to 2, give an abelian group law. That is, if $y^2 = f(x) = ax^3 + bx^2 + cx + d \in K[x]$ is the equation of an elliptic curve over K, and if we define $f'(x) = 3ax^2 + 2bx + c$, then any two points having coordinates in some extension of K can be added using the formulas (7.1)–(7.4). We shall make use of this fact in what follows, even though, strictly speaking, we have not gone through the tedious purely algebraic verification of the group laws.

PROBLEMS

1. Let $L \subset \mathbb{R}$ be the additive subgroup $\{m\omega\}$ of multiples of a fixed nonzero real number ω. Then the function $z \mapsto (\cos(2\pi z/\omega), \sin(2\pi z/\omega))$ gives a one-to-one analytic map of $\mathbb{R}/L$ onto the curve $x^2 + y^2 = 1$ in the real xy-plane. Show that ordinary addition in $\mathbb{R}/L$ carries over to a rational (actually polynomial) law for "adding" two points (x_1, y_1) and (x_2, y_2) on the unit circle; that is, the coordinates of the "sum" are polynomials in x_1, x_2, y_1, y_2. Thus, the rational addition law on an elliptic curve can be thought of as a generalization of the formulas for the sine and cosine of the sum of two angles.

2. (a) Simplify the expression for the x-coordinate of $2P$ in the case of the elliptic curve $y^2 = x^3 - n^2x$.
   (b) Let X, Y, Z be a rational right triangle with area n. Let P be the corresponding point on the curve $y^2 = x^3 - n^2x$ constructed in the text in §I.2. Let Q be the point constructed in Problem 2 of §I.2. Show that $P = 2Q$.
   (c) Prove that, if P is a point not of order 2 with rational coordinates on the curve $y^2 = x^3 - n^2x$, then the x-coordinate of $2P$ is the square of a rational number having even denominator. For example, the point $Q = ((41/7)^2, 720 \cdot 41/7^3)$ on the curve $y^2 = x^3 - 31^2x$ is not equal to twice a point P having rational coordinates. (In this problem, recall: n is always squarefree.)

3. Describe geometrically: (a) the four points of order two on an elliptic curve; (b) the nine points of order three; (c) how to find the twelve points of order four which are not of order two; (d) what the associative law of addition says about a certain configuration of lines joining points on the elliptic curve (draw a picture).

4. (a) How many points of inflection are there on an elliptic curve besides the point at infinity? Notice that they occur in symmetric pairs. Find an equation for their x-coordinates.
   (b) In the case of the elliptic curve $y^2 = x^3 - n^2x$ find an explicit formula for these x-coordinates. Show that they are never rational (for any n).

5. Given a point Q on an elliptic curve, how many points P are there such that $2P = Q$? Describe geometrically how to find them.

6. Show that if K is any subfield of $\mathbb{C}$ containing g_2 and g_3, then the points on the elliptic curve $y^2 = 4x^3 - g_2x - g_3$ whose coordinates are in K form a subgroup

of the group of all points. More generally, show that this is true for the elliptic curve $y^2 = f(x)$ if $f(x) \in K[x]$.

7. Consider the subgroup of all points on $y^2 = x^3 - n^2x$ with real coordinates. How many points in this subgroup are of order 2? 3? 4? Describe geometrically where these points are located.

8. Same as Problem 7 for the elliptic curve $y^2 = x^3 - a$, $a \in \mathbb{R}$.

9. If $y^2 = f(x)$ is an elliptic curve in which $f(x)$ has real coefficients, show that the group of points with real coordinates is isomorphic to (a) $\mathbb{R}/\mathbb{Z}$ if $f(x)$ has only one real root; (b) $\mathbb{R}/\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ if $f(x)$ has three real roots.

10. Letting a approach zero in Problem 8, show that for the curve $y^2 = x^3$ the same geometric procedure for finding $P_1 + P_2$ as for elliptic curves makes the smooth points of the curve (i.e., $P \neq (0, 0)$, but including the point at infinity) into an abelian group. Show that the map which takes $P = (x, y)$ to x/y (and takes the point at infinity to zero) gives an isomorphism with the additive group of complex numbers. This is called "additive degeneracy" of an elliptic curve. One way to think of this is to imagine both ω_1 and ω_2 approaching infinity (in different directions). Then g_2 and g_3 both approach zero, so the equation of the corresponding elliptic curve approaches $y^2 = 4x^3$. Meanwhile, the additive group $\mathbb{C}/L$, where $L = \{m\omega_1 + n\omega_2\}$, approaches the additive group $\mathbb{C}$, i.e., the fundamental parallelogram becomes all of $\mathbb{C}$.

11. Let $a \to 0$ in the elliptic curve $y^2 = (x^2 - a)(x + 1)$. Show that for the curve $y^2 = x^2(x + 1)$ the same geometric procedure for finding $P_1 + P_2$ as for elliptic curves makes the smooth points of the curve into an abelian group. Show that the map which takes $P = (x, y)$ to $(y - x)/(y + x)$ (and takes the point at infinity to 1) gives an isomorphism with the multiplicative group $\mathbb{C}^*$ of nonzero complex numbers. This is called "multiplicative degeneracy" of an elliptic curve. Draw the graph of the real points of $y^2 = x^2(x + 1)$, and show where the various sections go under the isomorphism with $\mathbb{C}^*$. One way to think of multiplicative degeneracy is to make the linear change of variables $y \mapsto \frac{i}{2}y$, $x \mapsto -x - \frac{1}{3}$, so that the equation becomes $y^2 = 4x^3 - \frac{4}{3}x - \frac{8}{27}$ (compare with Problem 8 of §I.6). So we are dealing with the limit as t approaches infinity of the group $\mathbb{C}/\{mit + n\pi\}$, i.e., with the vertical strip $\mathbb{C}/\{n\pi\}$ (rather, a cylinder, since opposite sides are glued together), and this is isomorphic to $\mathbb{C}^*$ under the map $z \mapsto e^{2iz}$.

§8. Points of finite order

In any group, there is a basic distinction between elements of finite order and elements of infinite order. In an abelian group, the set of elements of finite order form a subgroup, called the "torsion subgroup". In the case of the group of points in $\mathbb{P}^2_{\mathbb{C}}$ on the elliptic curve $y^2 = f(x)$, we immediately see that a point $P_z = (x, y)$ has finite order if and only if $Nz \in L$ for some N, i.e., if and only if z is a *rational* linear combination of ω_1 and ω_2. In that case, the least such N (which is the least common denominator of the coefficients of ω_1 and ω_2) is the exact order of P_z. Under the isomorphism

from $\mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$ to the elliptic curve given by $(a, b) \mapsto P_{a\omega_1 + b\omega_2}$, it is the image of $\mathbb{Q}/\mathbb{Z} \times \mathbb{Q}/\mathbb{Z}$ which is the torsion subgroup of the elliptic curve.

This situation is the two-dimensional analog of the circle group, whose torsion subgroup is precisely the group of all roots of unity, i.e., all $e^{2\pi i z}$ for $z \in \mathbb{Q}/\mathbb{Z}$. Just as the cyclotomic fields—the field extensions of $\mathbb{Q}$ generated by the roots of unity—are central to algebraic number theory, we would expect that the fields obtained by adjoining the coordinates of points $P = (x, y)$ of order N on an elliptic curve should have interesting special properties. We shall soon see that these coordinates are algebraic (if the coefficients of $f(x)$ are). This analogy between cyclotomic fields and fields formed from points of finite order on elliptic curves is actually much deeper than one might have guessed. In fact, a major area of research in algebraic number theory today consists in finding and proving analogs for such fields of the rich results one has for cyclotomic fields.

Let N be a fixed positive integer. Let $f(x) = ax^3 + bx^2 + cx + d = a(x - e_1)(x - e_2)(x - e_3)$ be a cubic polynomial with coefficients in a field K of characteristic $\neq 2$ and with distinct roots (perhaps in some extension of K). We are interested in describing the coordinates of the points of order N (i.e., exact order a divisor of N) on the elliptic curve $y^2 = f(x)$, where these coordinates may lie in an extension of K. If $N = 2$, the points of order N are the point at infinity 0 and $(e_i, 0)$, $i = 1, 2, 3$. Now suppose that $N > 2$. If N is odd, by a "nontrivial" point of order N we mean a point $P \neq 0$ such that $NP = 0$. If N is even, by a "nontrivial" point of order N we mean a point P such that $NP = 0$ but $2P \neq 0$.

Proposition 13. *Let K' be any field extension of K (not necessarily algebraic), and let $\sigma\colon K' \to \sigma K'$ be any field isomorphism which leaves fixed all elements of K. Let $P \in \mathbb{P}^2_{K'}$ be a point of exact order N on the elliptic curve $y^2 = f(x)$, where $f(x) \in K[x]$. Then σP has exact order N (where for $P = (x, y, z) \in \mathbb{P}^2_{K'}$ we denote $\sigma P \underset{\text{def}}{=} (\sigma x, \sigma y, \sigma z) \in \mathbb{P}^2_{\sigma K'}$).*

Proof. It follows from the addition formulas that $\sigma P_1 + \sigma P_2 = \sigma(P_1 + P_2)$, and hence $N(\sigma P) = \sigma(NP) = \sigma 0 = 0$ (since $\sigma(0, 1, 0) = (0, 1, 0)$). Hence σP has order N. It must have exact order N, since if $N'\sigma P = 0$, we would have $\sigma(N'P) = 0 = (0, 1, 0)$, and hence $N'P = 0$. This proves the proposition. □

Proposition 14. *In the situation of Proposition 13, with K a subfield of $\mathbb{C}$, let $K_N \subset \mathbb{C}$ denote the field obtained by adjoining to K the x- and y-coordinates of all points of order N. Let K_N^+ denote the field obtained by adjoining just their x-coordinates. Then both K_N and K_N^+ are finite galois extensions of K.*

Proof. In each case K_N and K_N^+, we are adjoining a finite set of complex numbers which are permuted by any automorphism of $\mathbb{C}$ which fixes K. This immediately implies the proposition. □

As an example, if $N = 2$, then $K_2 = K_2^+$ is the splitting field of $f(x)$ over K.

Recall that the group of points of order N on an elliptic curve in $\mathbb{P}^2_{\mathbb{C}}$ is isomorphic to $(\mathbb{Z}/N\mathbb{Z}) \times (\mathbb{Z}/N\mathbb{Z})$. Because any $\sigma \in \mathrm{Gal}(K_N/K)$ respects addition of points, i.e., $\sigma(P_1 + P_2) = \sigma P_1 + \sigma P_2$, it follows that each σ gives an invertible linear map of $(\mathbb{Z}/N\mathbb{Z})^2$ to itself.

If R is any commutative ring, we let $GL_n(R)$ denote the group (under matrix multiplication) of all $n \times n$ invertible matrices with entries in R. Here invertibility of a matrix A is equivalent to $\det A \in R^*$, where R^* is the multiplicative group of invertible elements of the ring. For example:

(1) $GL_1(R) = R^*$;
(2) $GL_2(\mathbb{Z}/N\mathbb{Z}) = \{(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}) | a, b, c, d \in \mathbb{Z}/N\mathbb{Z}, ad - bc \in (\mathbb{Z}/N\mathbb{Z})^*\}$.

It is easy to construct a natural one-to-one correspondence between invertible linear maps $R^n \to R^n$ and elements of $GL_n(R)$. There is no difference with the more familiar case when R is a field.

In our situation of points of order N on an elliptic curve, we have seen that $\mathrm{Gal}(K_N/K)$ is isomorphic to a subgroup of the group of all invertible linear maps $(\mathbb{Z}/N\mathbb{Z})^2 \to (\mathbb{Z}/N\mathbb{Z})^2$. Thus, any $\sigma \in \mathrm{Gal}(K_N/K)$ corresponds to a matrix $(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}) \in GL_2(\mathbb{Z}/N\mathbb{Z})$. The matrix entries can be found by writing

$$\sigma P_{\omega_1/N} = P_{a\omega_1/N + c\omega_2/N}, \qquad \sigma P_{\omega_2/N} = P_{b\omega_1/N + d\omega_2/N}.$$

Notice that this is a direct generalization of the situation with the N-th cyclotomic field $\mathbb{Q}_N \underset{\text{def}}{=} \mathbb{Q}(\sqrt[N]{1})$. Recall that $\mathrm{Gal}(\mathbb{Q}_N/\mathbb{Q}) \approx (\mathbb{Z}/N\mathbb{Z})^* = GL_1(\mathbb{Z}/N\mathbb{Z})$, with the element a which corresponds to σ determined by

$$\sigma(e^{2\pi i/N}) = e^{2\pi i a/N}.$$

But one difference in our two-dimensional case of division points on elliptic curves is that, in general, $\mathrm{Gal}(K_N/K) \to GL_2(\mathbb{Z}/N\mathbb{Z})$ is only an injection, not an isomorphism.

In the case $K \subset \mathbb{C}$, say $K = \mathbb{Q}(g_2, g_3)$, where $y^2 = f(x) = 4x^3 - g_2 x - g_3$ is in Weierstrass form, we shall now use the $\wp$-function to determine the polynomial whose roots are the x-coordinates of the points of order N. That is, K_N^+ will be the splitting field of such a polynomial.

We first construct an elliptic function $f_N(z)$ whose zeros are precisely the nonzero values of z such that P_z is a point of order N. We follow the prescription in the proof of Proposition 9 of §I.5. If $u \in \mathbb{C}/L$ is a point of order N, then so is the symmetric point $-u$ (which we denoted u^* when we were thinking in terms of points in a fundamental parallelogram). We consider two cases:

(i) N is odd. Then the points u and $-u$ are always distinct modulo L. In other words, u cannot be $\omega_1/2$, $\omega_2/2$ or $(\omega_1 + \omega_2)/2$ if u has odd order N. We define

$$f_N(z) = N \prod (\wp(z) - \wp(u)), \tag{8.1}$$

where the product is taken over nonzero $u \in \mathbb{C}/L$ such that $Nu \in L$, with one u taken from each pair $u, -u$. Then $f_N(z) = F_N(\wp(z))$, where $F_N(x) \in \mathbb{C}[x]$ is a polynomial of degree $(N^2-1)/2$. The even elliptic function $f_N(z)$ has N^2-1 simple zeros and a single pole at 0 of order N^2-1. Its leading term at $z=0$ is N/z^{N^2-1}.

(ii) N is even. Now let u range over $u \in \mathbb{C}/L$ such that $Nu \in L$ but u is *not* of order 2, i.e., $u \neq 0, \omega_1/2, \omega_2/2, (\omega_1+\omega_2)/2$. Define $\tilde{f}_N(z)$ by the product in (8.1). Then $\tilde{f}_N(z) = F_N(\wp(z))$, where $F_N(x) \in \mathbb{C}[x]$ is a polynomial of degree $(N^2-4)/2$. The even elliptic function $\tilde{f}_N(z)$ has N^2-4 simple zeros and a single pole at 0 of order N^2-4. Its leading term at $z=0$ is N/z^{N^2-4}.

If N is odd, the function $f_N(z)$ has the property that

$$f_N(z)^2 = N^2 \prod_{0 \neq u \in \mathbb{C}/L,\, Nu \in L} (\wp(z) - \wp(u)).$$

If N is even, then the function $f_N(z) \underset{\text{def}}{=} -\frac{1}{2}\wp'(z)\tilde{f}_N(z)$ has the property that

$$\begin{aligned} f_N(z)^2 &= \tfrac{1}{4}\wp'(z)^2\tilde{f}_N(z)^2 \\ &= N^2(\wp(z)-e_1)(\wp(z)-e_2)(\wp(z)-e_3) \prod_{u \in \mathbb{C}/L,\, Nu \in L,\, 2u \notin L} (\wp(z)-\wp(u)) \\ &= N^2 \prod_{0 \neq u \in \mathbb{C}/L,\, Nu \in L} (\wp(z)-\wp(u)). \end{aligned}$$

We see that a point $(x, y) = (\wp(z), \wp'(z))$ has odd order N if and only if $F_N(x) = 0$. It has even order N if and only if either $y = 0$ (i.e., it is a point of order 2) or else $F_N(x) = 0$.

Because of Propositions 13 and 14, we know that any automorphism of $\mathbb{C}$ fixing $K = \mathbb{Q}(g_2, g_3)$ permutes the roots of F_N. Hence, the coefficients of F_N are in $K = \mathbb{Q}(g_2, g_3)$.

If we started with an elliptic curve not in Weierstrass form, say $y^2 = f(x) = ax^3 + bx^2 + cx + d$, and if we wanted to avoid using the $\wp$-function, then we could repeatedly apply the addition formulas (7.1)–(7.4) to compute the rational function of x and y which is the x-coordinate of NP, where $P = (x, y)$. We would simplify algebraically as we go, making use of the relation $y^2 = f(x)$, and would end up with an expression in the denominator which vanishes if and only if NP is the point at infinity, i.e., if and only if P has order N (recall: "order N" means "exact order N or a divisor of N").

What type of an expression would we have to get in the denominator of the x-coordinate of NP? Suppose, for example, that N is odd. Then this denominator would be an expression in $K[x, y]$ (with y occurring at most to the first power), where $K = \mathbb{Q}(a, b, c, d)$, which vanishes if and only if x is one of the $(N^2-1)/2$ values of x-coordinates of nontrivial points of order N. Thus, the expression must be a polynomial in x alone with $(N^2-1)/2$ roots. Similarly, we find that when N is even, this denominator has the form

$y \cdot$(polynomial in x alone), where the polynomial in $K[x]$ has $(N^2 - 4)/2$ roots.

It is important to note that the algebraic procedure described in the last two paragraphs applies for any elliptic curve $y^2 = f(x)$ over any field K of characteristic $\neq 2$, not only over subfields of the complex numbers. Thus, for any K we end up with an expression in the denominator of the x-coordinate of NP that vanishes for at most $N^2 - 1$ values of (x, y).

For a general field K, however, we do not necessarily get exactly $N^2 - 1$ nontrivial points of order N. Of course, if K is not algebraically closed, the coordinates of points of order N may lie only in some extension of K. Moreover, if K has characteristic p, then there might be fewer points of order N for another reason: the leading coefficient of the expression in the denominator vanishes modulo p, and so the degree of that polynomial drops. We shall soon see examples where there are fewer than N^2 points of order N even if we allow coordinates in $K^{\text{alg cl}}$.

This discussion has led to the following proposition.

Proposition 15. *Let $y^2 = f(x)$ be an elliptic curve over any field K of characteristic not equal to 2. Then there are at most N^2 points of order N over any extension K' of K.*

Now let us turn our attention briefly to the case of K a finite field, in order to illustrate one application of Proposition 15. We shall later return to elliptic curves over finite fields in more detail.

Since there are only finitely many points in $\mathbb{P}^2_{\mathbb{F}_q}$ (namely, $q^2 + q + 1$), there are certainly only finitely many $\mathbb{F}_q$-points on an elliptic curve $y^2 = f(x)$, where $f(x) \in \mathbb{F}_q[x]$. So the group of $\mathbb{F}_q$-points is a *finite abelian group*.

By the "type" of a finite abelian group, we mean its expression as a product of cyclic groups of prime power order. We list the orders of all of the cyclic groups that appear in the form: $2^{\alpha_2}, 2^{\beta_2}, 2^{\gamma_2}, \ldots, 3^{\alpha_3}, 3^{\beta_3}, 3^{\gamma_3}, \ldots, 5^{\alpha_5}, 5^{\beta_5}, \ldots$. But Proposition 15 implies that only certain types can occur in the case of the group of $\mathbb{F}_q$-points on $y^2 = f(x)$. Namely, for each prime l there are at most two l-th power components $l^{\alpha_l}, l^{\beta_l}$, since otherwise we would have more than l^2 points of order l. And of course $l^{\alpha_l + \beta_l}$ must equal the power of l dividing the order of the group.

As an example of how this works, let us consider the elliptic curve $y^2 = x^3 - n^2 x$ over $K = \mathbb{F}_q$ (the finite field of $q = p^f$ elements), where we must assume that p does not divide $2n$. In the case when $q \equiv 3 \pmod 4$, it is particularly easy to count the number of $\mathbb{F}_q$-points.

Proposition 16. *Let $q = p^f$, $p \nmid 2n$. Suppose that $q \equiv 3 \pmod 4$. Then there are $q + 1$ $\mathbb{F}_q$-points on the elliptic curve $y^2 = x^3 - n^2 x$.*

Proof. First, there are four points of order 2: the point at infinity, $(0, 0)$, and $(\pm n, 0)$. We now count all pairs (x, y) where $x \neq 0, n, -n$. We arrange

these $q-3$ x's in pairs $\{x, -x\}$. Since $f(x) = x^3 - n^2x$ is an odd function, and -1 is not a square in $\mathbb{F}_q$ (here's where we use the assumption that $q \equiv 3 \pmod 4$), it follows that exactly one of the two elements $f(x)$ and $f(-x) = -f(x)$ is a square in $\mathbb{F}_q$. (Recall: In the multiplicative group of a finite field, the squares are a subgroup of index 2, and so the product of two nonsquares is a square, while the product of a square and a nonsquare is a nonsquare.) Whichever of the pair x, $-x$ gives a square, we obtain exactly two points $(x, \pm\sqrt{f(x)})$ or else $(-x, \pm\sqrt{f(-x)})$. Thus, the $(q-3)/2$ pairs give us $q-3$ points. Along with the four points of order two, we have $q+1$ $\mathbb{F}_q$-points in all, as claimed. □

Notice that when $q \equiv 3 \pmod 4$, the number of $\mathbb{F}_q$-points on the elliptic curve $y^2 = x^3 - n^2x$ does not depend on n. This is not true if $q \equiv 1 \pmod 4$.

As an example, Proposition 16 tells us that for $q = 7^3$ there are $344 = 2^3 \cdot 43$ points. Since there are four points of order two, the type of the group of $\mathbb{F}_{343}$-points on $y^2 = x^3 - n^2x$ must be $(2, 2^2, 43)$.

As a more interesting example, let $q = p = 107$. Then there are $108 = 2^2 \cdot 3^3$ points. The group is either of type $(2, 2, 3^3)$ or of type $(2, 2, 3, 3^2)$. To resolve the question, we must determine whether there are 3 or 9 points of order three. (There must be nontrivial points of order 3, since 3 divides the order of the group.) Recall the equation for the x-coordinates of points of order three (see Problem 4 of §7): $-3x^4 + 6n^2x^2 + n^4 = 0$, i.e., $x = \pm n\sqrt{1 \pm 2\sqrt{3}/3}$. Then the corresponding y-coordinates are found by taking $\pm\sqrt{f(x)}$. We want to know how many of these points have both coordinates in $\mathbb{F}_{107}$, rather than an extension of $\mathbb{F}_{107}$. We could compute explicitly, using $\sqrt{3} = \pm 18$ in $\mathbb{F}_{107}$, so that $x = \pm\sqrt{13}, \pm\sqrt{-11}$, etc. But even before doing those computations, we can see that not all 9 points have coordinates in $\mathbb{F}_{107}$. This is because, if (x, y) is in $\mathbb{F}_{107}$, then $(-x, \sqrt{-1}y)$ is another point of order three, and its coordinates are not in $\mathbb{F}_{107}$. Thus, there are only 3 points of order three, and the type of the group is $(2, 2, 3^3)$.

Notice that if K is any field of characteristic 3, then the group of K-points has *no* nontrivial point of order three, because $-3x^4 + 6n^2x^2 + n^4 = n^4 \neq 0$. This is an example of the "dropping degree" phenomenon mentioned above. It turns out that the same is true for any $p \equiv 3 \pmod 4$, namely, there are no points of order p over a field of characteristic p in that case. This is related to the fact that such p remain prime in the ring of Gaussian integers $\mathbb{Z}[i]$, a ring which is intimately related to our particular elliptic curve (see Problem 13 of §6). But we will not go further into that now.

Problems

1. For the elliptic curve $y^2 = 4x^3 - g_2x - g_3$, express $\wp(Nz)$ as a rational function of $\wp(z)$ when $N = 2$.

2. Let $f_N(z)$ be the elliptic functions defined above. Express $f_3(z)$ as a polynomial in $\wp(z)$.

3. Set $f_1(z) = 1$. Prove that for $N = 2, 3, 4, \ldots$ we have:

$$\wp(Nz) = \wp(z) - f_{N-1}(z)f_{N+1}(z)/f_N(z)^2.$$

4. In the notation of Proposition 14, suppose that $\sigma \in \mathrm{Gal}(K_N/K)$ fixes all x-coordinates of points of order N. That is, $\sigma|_{K_N^+} = $ identity. Show that the image of σ in $GL_2(\mathbb{Z}/N\mathbb{Z})$ is ± 1. Conclude that $\mathrm{Gal}(K_N/K_N^+) = \{\pm 1\} \cap G$, where G is the image of $\mathrm{Gal}(K_N/K)$ in $GL_2(\mathbb{Z}/N\mathbb{Z})$. What is the analogous situation for cyclotomic fields?

5. Let $L = \{m\omega_1 + n\omega_2\}$, and let E be the elliptic curve $y^2 = 4x^3 - g_2(L)x - g_3(L)$. Notice that E does not change if we replace the basis $\{\omega_1, \omega_2\}$ of L by another basis $\{\omega_1', \omega_2'\}$. However, the group isomorphism $\mathbb{C}/L \approx \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$ changes, and so does the isomorphism from the points of order N on E to $\mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$. For example, the point $(\wp(\omega_1'/N), \wp'(\omega_1'/N))$, rather than $(\wp(\omega_1/N), \wp'(\omega_1/N))$, corresponds to $(1, 0) \in \mathbb{Z}/N\mathbb{Z} \times \mathbb{Z}/N\mathbb{Z}$. What effect does the change of basis from ω_i to ω_i' have on the image of $\mathrm{Gal}(K_N/K)$ in $GL_2(\mathbb{Z}/N\mathbb{Z})$?

6. Show that the group $GL_2(\mathbb{Z}/2\mathbb{Z})$ is isomorphic to S_3, the group of permutations of $\{1, 2, 3\}$. For each of the following elliptic curves, describe the image in $GL_2(\mathbb{Z}/2\mathbb{Z})$ of the galois group over $\mathbb{Q}$ of the field generated by the coordinates of the points of order 2.
   (a) $y^2 = x^3 - nx$ (n not a perfect square)
   (b) $y^2 = x^3 - n^2x$
   (c) $y^2 = x^3 - n$ (n not a perfect cube)
   (d) $y^2 = x^3 - n^3$.

7. (a) How many elements are in $GL_2(\mathbb{Z}/3\mathbb{Z})$?
   (b) Describe the field extension K_3 of $K = \mathbb{Q}$ generated by the coordinates of all points of order 3 on the elliptic curve $y^2 = x^3 - n^2x$.
   (c) Find $[K_3 : \mathbb{Q}]$. What subgroup of $GL_2(\mathbb{Z}/3\mathbb{Z})$ is isomorphic to $\mathrm{Gal}(K_3/\mathbb{Q})$?
   (d) Give a simple example of an element in $GL_2(\mathbb{Z}/3\mathbb{Z})$ that is *not* in the image of $\mathrm{Gal}(K_3/\mathbb{Q})$; in other words, find a pair of elements $z_1 = (m_1\omega_1 + n_1\omega_2)/3$, $z_2 = (m_2\omega_1 + n_2\omega_2)/3$ which generate all $(m\omega_1 + n\omega_2)/3$ but such that P_{z_1}, P_{z_2} cannot be obtained from $P_{\omega_1/3}$, $P_{\omega_2/3}$ by applying an automorphism to the coordinates of the latter pair of points.

8. In Problem 13 of §I.6, we saw that the lattice corresponding to the curve $y^2 = x^3 - n^2x$ is the lattice L of Gaussian integers expanded by a factor $\omega_2 \in \mathbb{R}$: $L = \{mi\omega_2 + n\omega_2\} = \omega_2\mathbb{Z}[i]$.
   (a) Show that the map $z \mapsto iz$ gives an analytic automorphism of the additive group $\mathbb{C}/L$; and, more generally, for any Gaussian integer $a + bi \in \mathbb{Z}[i]$ we have a corresponding analytic endomorphism of $\mathbb{C}/L$ induced by $z \mapsto (a + bi)z$.
   (b) Notice that if $b = 0$, this is the map $z \mapsto z + z + \cdots + z$ (a times) which gives ϕ_a: $P \mapsto aP$ on the elliptic curve. By looking at the definition of $\wp(z)$, $\wp'(z)$, show that the map $z \mapsto iz$ gives the automorphism ϕ_i: $(x, y) \mapsto (-x, iy)$ on the elliptic curve. This is an example of what's called "complex multiplication". Show that $\phi_i \circ \phi_i = \phi_{-1}$, and in fact the map $a + bi \mapsto \phi_{a+bi}$ is an injection of the ring $\mathbb{Z}[i]$ into the ring of endomorphisms of the group of points on the elliptic curve.
   (c) If L is a lattice in $\mathbb{C}$ and if there exists a complex number $\alpha = a + bi$, $b \neq 0$, such that $\alpha L \subset L$, show that α is a quadratic imaginary algebraic integer, and that L contains a sublattice of finite index of the form $\omega_2\mathbb{Z}[\alpha]$.

9. Each of the following points has finite order N on the given elliptic curve. In each case, find its order.
   (a) $P = (0, 4)$ on $y^2 = 4x^3 + 16$
   (b) $P = (2, 8)$ on $y^2 = 4x^3 + 16x$
   (c) $P = (2, 3)$ on $y^2 = x^3 + 1$
   (d) $P = (3, 8)$ on $y^2 = x^3 - 43x + 166$
   (e) $P = (3, 12)$ on $y^2 = x^3 - 14x^2 + 81x$
   (f) $P = (0, 0)$ on $y^2 + y = x^3 - x^2$
   (g) $P = (1, 0)$ on $y^2 + xy + y = x^3 - x^2 - 3x + 3$.

§9. Points over finite fields, and the congruent number problem

We have mainly been interested in elliptic curves E over $\mathbb{Q}$, particularly the elliptic curve $y^2 = x^3 - n^2x$, which we shall denote E_n. But if K is any field whose characteristic p does not divide $2n$, the same equation (where we consider n modulo p) is an elliptic curve over K. We shall let $E_n(K)$ denote the set of points on the curve with coordinates in K. Thus, Proposition 16 in the last section can be stated: If $q \equiv 3 \pmod 4$, then $\# E_n(\mathbb{F}_q) = q + 1$.

The elliptic curve E_n considered as being defined over $\mathbb{F}_p$, is called the "reduction" modulo p, and we say that E_n has "good reduction" if p does not divide $2n$, i.e., if $y^2 = x^3 - n^2x$ gives an elliptic curve over $\mathbb{F}_p$. More generally, if $y^2 = f(x)$ is an elliptic curve E defined over an algebraic number field, and if $\mathfrak{p}$ is a prime ideal of the number field which does not divide the denominators of the coefficients of $f(x)$ or the discriminant of $f(x)$, then by reduction modulo $\mathfrak{p}$ we obtain an elliptic curve defined over the (finite) residue field of $\mathfrak{p}$.

At first glance, it may seem that the elliptic curves over finite fields—which lead only to finite abelian groups—are not a serious business, and that reduction modulo p is a frivolous game that will not help us in our original objective of studying $\mathbb{Q}$-points on $y^2 = x^3 - n^2x$. However, this is far from the case. Often information from the various reductions modulo p can be pieced together to yield information about the $\mathbb{Q}$-points. This is usually a subtle and difficult procedure, replete with conjectures and unsolved problems. However, there is one result of this type which is simple enough to give right now. Namely, we shall use reduction modulo p for various primes p to determine the torsion subgroup of $E_n(\mathbb{Q})$, the group of $\mathbb{Q}$-points on $y^2 = x^3 - n^2x$.

In any abelian group, the elements of finite order form a subgroup, called the "torsion subgroup". For example, the group $E(\mathbb{C})$ of complex points on an elliptic curve is isomorphic to $\mathbb{C}/L$, which for any lattice L is isomorphic to $\mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$ (see Problem 2 of §I.5). Its torsion subgroup corresponds to the subgroup $\mathbb{Q}/\mathbb{Z} \times \mathbb{Q}/\mathbb{Z} \subset \mathbb{R}/\mathbb{Z} \times \mathbb{R}/\mathbb{Z}$, i.e., in $\mathbb{C}/L$ it consists of all rational linear combinations of ω_1 and ω_2.

A basic theorem of Mordell states that the group $E(\mathbb{Q})$ of $\mathbb{Q}$-points on an

elliptic curve E defined over $\mathbb{Q}$ is a finitely generated abelian group. This means that (1) the torsion subgroup $E(\mathbb{Q})_{\text{tors}}$ is finite, and (2) $E(\mathbb{Q})$ is isomorphic to the direct sum of $E(\mathbb{Q})_{\text{tors}}$ and a finite number of copies of $\mathbb{Z}$: $E(\mathbb{Q}) \approx E(\mathbb{Q})_{\text{tors}} \oplus \mathbb{Z}^r$. The nonnegative integer r is called the "rank" of $E(\mathbb{Q})$. It is greater than zero if and only if E has infinitely many $\mathbb{Q}$-points. Mordell's theorem is also true, by the way, if $\mathbb{Q}$ is replaced by any algebraic number field. This generalization, proved by André Weil, is known as the Mordell–Weil theorem. We shall not need this theorem for our purposes, even in the form proved by Mordell. For a proof, the reader is referred to [Husemöller 1987] or [Lang 1978b].

We shall now prove that the only rational points of finite order on E_n are the four points of order 2: 0 (the point at infinity), (0, 0), $(\pm n, 0)$.

Proposition 17. $\# E_n(\mathbb{Q})_{\text{tors}} = 4$.

PROOF. The idea of the proof is to construct a group homomorphism from $E_n(\mathbb{Q})_{\text{tors}}$ to $E_n(\mathbb{F}_p)$ which is injective for most p. That will imply that the order of $E_n(\mathbb{Q})_{\text{tors}}$ divides the order of $E_n(\mathbb{F}_p)$ for such p. But no number greater than 4 could divide all such numbers $\# E_n(\mathbb{F}_p)$, because we at least know that $\# E_n(\mathbb{F}_p)$ runs through all integers of the form $p + 1$ for p a prime congruent to 3 modulo 4 (see Proposition 16).

We begin the proof of Proposition 17 by constructing the homomorphism from the group of $\mathbb{Q}$-points on E_n to the group of $\mathbb{F}_p$-points. More generally, we simply construct a map from $\mathbb{P}^2_{\mathbb{Q}}$ to $\mathbb{P}^2_{\mathbb{F}_p}$. In what follows, we shall always choose a triple (x, y, z) for a point in $\mathbb{P}^2_{\mathbb{Q}}$ in such a way that x, y, and z are integers with no common factor. Up to multiplication by ± 1, there is a unique such triple in the equivalence class. For any fixed prime p, we define the image $\bar{P}$ of $P = (x, y, z) \in \mathbb{P}^2_{\mathbb{Q}}$ to be the point $\bar{P} = (\bar{x}, \bar{y}, \bar{z}) \in \mathbb{P}^2_{\mathbb{F}_p}$, where the bar denotes reduction of an integer modulo p. Note that $\bar{P}$ is not the identically zero triple, because p does not divide all three integers x, y, z. Also note that we could have replaced the triple (x, y, z) by any multiple by an integer prime to p without affecting $\bar{P}$.

It is easy to see that if $P = (x, y, z)$ happens to be in $E_n(\mathbb{Q})$, i.e., if $y^2z = x^3 - n^2xz^2$, then $\bar{P}$ is in $E_n(\mathbb{F}_p)$. Moreover, the image of $P_1 + P_2$ under this map is $\bar{P}_1 + \bar{P}_2$, because it makes no difference whether we use the addition formulas (7.1)–(7.4) to find the sum and then reduce mod p, or whether we first reduce mod p and then use the addition formulas. In other words, our map is a homomorphism from $E_n(\mathbb{Q})$ to $E_n(\mathbb{F}_p)$, for any prime p not dividing $2n$.

We now determine when this map is not injective, i.e., when two points $P_1 = (x_1, y_1, z_1)$ and $P_2 = (x_2, y_2, z_2)$ in $\mathbb{P}^2_{\mathbb{Q}}$ have the same image $\bar{P}_1 = \bar{P}_2$ in $\mathbb{P}^2_{\mathbb{F}_p}$.

Lemma. $\bar{P}_1 = \bar{P}_2$ *if and only if the cross-product of* P_1 *and* P_2 *(considered as vectors in* $\mathbb{R}^3$*) is divisible by* p*, i.e., if and only if* p *divides* $y_1z_2 - y_2z_1$, $x_2z_1 - x_1z_2$, *and* $x_1y_2 - x_2y_1$.

PROOF OF LEMMA. First suppose that p divides the cross-product. We consider two cases:

(i) p divides x_1. Then p divides $x_2 z_1$ and $x_2 y_1$, and therefore divides x_2, because it cannot divide x_1, y_1 and z_1. Suppose, for example, that $p \nmid y_1$ (an analogous argument will apply if $p \nmid z_1$). Then $\bar{P}_2 = (0, \bar{y}_1\bar{y}_2, \bar{y}_1\bar{z}_2) = (0, \bar{y}_1\bar{y}_2, \bar{y}_2\bar{z}_1) = (0, \bar{y}_1, \bar{z}_1) = \bar{P}_1$ (where we have used the fact that p divides $y_1 z_2 - y_2 z_1$).

(ii) p does not divide x_1. Then $\bar{P}_2 = (\bar{x}_1\bar{x}_2, \bar{x}_1\bar{y}_2, \bar{x}_1\bar{z}_2) = (\bar{x}_1\bar{x}_2, \bar{x}_2\bar{y}_1, \bar{x}_2\bar{z}_1) = (\bar{x}_1, \bar{y}_1, \bar{z}_1) = \bar{P}_1$.

Conversely, suppose that $\bar{P}_1 = \bar{P}_2$. Without loss of generality, suppose that $p \nmid x_1$ (an analogous argument will apply if $p \nmid y_1$ or $p \nmid z_1$). Then, since $\bar{P}_1 = \bar{P}_2 = (\bar{x}_2, \bar{y}_2, \bar{z}_2)$, we also have $p \nmid x_2$. Hence, $(\bar{x}_1\bar{x}_2, \bar{x}_1\bar{y}_2, \bar{x}_1\bar{z}_2) = \bar{P}_2 = \bar{P}_1 = (\bar{x}_2\bar{x}_1, \bar{x}_2\bar{y}_1, \bar{x}_2\bar{z}_1)$. Since the first coordinates are the same, these two points can be equal only if the second and third coordinates are equal, i.e., if p divides $x_1 y_2 - x_2 y_1$ and $x_1 z_2 - x_2 z_1$. Finally, we must show that p divides $y_1 z_2 - y_2 z_1$. If both y_1 and z_1 are divisible by p, then this is trivial. Otherwise, the conclusion will follow by repeating the above argument with x_1, x_2 replaced by y_1, y_2 or by z_1, z_2. This concludes the proof of the lemma.

We are now ready to prove Proposition 17. Suppose that the proposition is false, i.e., that $E_n(\mathbb{Q})$ contains a point of finite order greater than 2. Then either it contains an element of odd order, or else the group of points of order 4 (or a divisor of 4) contains either 8 or 16 elements. In either case we have a subgroup $S = \{P_1, P_2, \ldots, P_m\} \subset E_n(\mathbb{Q})_{\text{tors}}$, where $m = \#S$ is either 8 or else an odd number.

Let us write all of the points P_i, $i = 1, \ldots, m$, in the form in the lemma: $P_i = (x_i, y_i, z_i)$. For each pair of points P_i, P_j, consider the cross-product vector $(y_i z_j - y_j z_i, x_j z_i - x_i z_j, x_i y_j - x_j y_i) \in \mathbb{R}^3$. Since P_i and P_j are distinct points, as vectors in $\mathbb{R}^3$ they are not proportional, and so their cross-product is not the zero vector. Let n_{ij} be the greatest common divisor of the coordinates of this cross-product. According to the lemma, the points P_i and P_j have the same image $\bar{P}_i = \bar{P}_j$ in $E_n(\mathbb{F}_p)$ if and only if p divides n_{ij}. Thus, if p is a prime of good reduction which is greater than all of the n_{ij}, it follows that all images are distinct, i.e., the map reduction modulo p gives an *injection* of S in $E_n(\mathbb{F}_p)$.

But this means that for all but finitely many p the number m must divide $\#E_n(\mathbb{F}_p)$, because the image of S is a subgroup of order m. Then for all but finitely many primes congruent to 3 modulo 4, by Proposition 16 we must have $p \equiv -1 \pmod{m}$. But this contradicts Dirichlet's theorem on primes in an arithmetic progression. Namely, if $m = 8$ this would mean that there are only finitely many primes of the form $8k + 3$. If m is odd, it would mean that there are only finitely many primes of the form $4mk + 3$ (if $3 \nmid m$), and that there are only finitely many primes of the form $12k + 7$ if $3|m$. In all cases, Dirichlet's theorem tells us that there are infinitely many primes of the given type. This concludes the proof of Proposition 17. □

Notice how the technique of reduction modulo p (more precisely, the use of Proposition 16 for infinitely many primes p) led to a rather painless proof of a strong fact: There are no "non-obvious" rational points of finite order on E_n. As we shall soon see, this fact is useful for the congruent number problem. But a far more interesting and difficult question is the existence of points of infinite order, i.e., whether the rank r of $E_n(\mathbb{Q})$ is nonzero. As we shall see in a moment, that question is actually *equivalent* to the question of whether or not n is a congruent number.

So it is natural to ask whether mod p information can somehow be put together to yield information about the rank of an elliptic curve. This subtle question will lead us in later chapters to consideration of the Birch–Swinnerton-Dyer conjecture for elliptic curves.

For further general motivational discussion of elliptic curves over finite fields, see [Koblitz 1982].

We now prove the promised corollary of Proposition 17.

Proposition 18. *n is a congruent number if and only if $E_n(\mathbb{Q})$ has nonzero rank r.*

Proof. First suppose that n is a congruent number. At the beginning of §2, we saw that the existence of a right triangle with rational sides and area n leads to a rational point on E_n whose x-coordinate lies in $(\mathbb{Q}^+)^2$. Since the x-coordinates of the three nontrivial points of order 2 are 0, $\pm n$, this means that there must be a rational point not of order 2. By Proposition 17, such a point has infinite order, i.e., $r \geq 1$.

Conversely, suppose that P is a point of infinite order. By Problem 2(c) of §I.7, the x-coordinate of the point $2P$ is the square of a rational number having even denominator. Now by Proposition 2 in §I.2, the point $2P$ corresponds to a right triangle with rational sides and area n (under the correspondence in Proposition 1). This proves Proposition 18. □

Notice the role of Proposition 17 in the proof of Proposition 18. It tells us that the only way to get nontrivial rational points of the form $2P$ is from points of infinite order. Let $2E_n(\mathbb{Q})$ denote the subgroup of $E_n(\mathbb{Q})$ consisting of the doubles of rational points. Then Proposition 17 is equivalent to the assertion that $2E_n(\mathbb{Q})$ is a torsion-free abelian group, i.e., it is isomorphic to a certain number of copies (namely, r) of $\mathbb{Z}$. The set $2E_n(\mathbb{Q}) - 0$ (0 denotes the point at infinity) is empty if and only if $r = 0$.

We saw that points in the set $2E_n(\mathbb{Q}) - 0$ lead to right triangles with rational sides and area n under the correspondence in Proposition 1. It is natural to ask whether all points meeting the conditions in Proposition 2, i.e., corresponding to triangles, are doubles of points. We now prove that the answer is yes. At the same time, we give another verification of Proposition 18 (not relying on the homework problem 2(c) of §I.7).

Proposition 19. *There is a one-to-one correspondence between right triangles with rational sides $X < Y < Z$ and area n, and pairs of points $(x, \pm y) \in$*

$2E_n(\mathbb{Q}) - 0$. *The correspondence is*:

$$(x, \pm y) \mapsto \sqrt{x+n} - \sqrt{x-n}, \sqrt{x+n} + \sqrt{x-n}, 2\sqrt{x};$$
$$X, Y, Z \mapsto (Z^2/4, \pm(Y^2 - X^2)Z/8).$$

In light of Proposition 1 of §I.1, Proposition 19 is an immediate consequence of the following general characterization of the doubles of points on elliptic curves.

Proposition 20. *Let E be the elliptic curve $y^2 = (x - e_1)(x - e_2)(x - e_3)$ with $e_1, e_2, e_3 \in \mathbb{Q}$. Let $P = (x_0, y_0) \in E(\mathbb{Q}) - 0$. Then $P \in 2E(\mathbb{Q}) - 0$ if and only if $x_0 - e_1, x_0 - e_2, x_0 - e_3$ are all squares of rational numbers.*

Proof. We first note that, without loss of generality, we may assume that $x_0 = 0$. To see this, make the change of variables $x' = x - x_0$. By simply translating the geometrical picture for adding points, we see that the point $P' = (0, y_0)$ on the curve E' with equation $y^2 = (x - e_1')(x - e_2')(x - e_3')$, where $e_i' = e_i - x_0$, is in $2E'(\mathbb{Q}) - 0$ if and only if our original P were in $2E(\mathbb{Q}) - 0$. And trivially, the $x_0 - e_i$ are all squares if and only if the $(0 - e_i')$ are. So it suffices to prove the proposition with $x_0 = 0$.

Next, note that if there exists $Q \in E(\mathbb{Q})$ such that $2Q = P$, then there are exactly four such points $Q, Q_1, Q_2, Q_3 \in E(\mathbb{Q})$ with $2Q_i = P$. To obtain Q_i, simply add to Q the point of order two $(e_i, 0) \in E(\mathbb{Q})$ (see Problem 5 in §I.7).

Choose a point $Q = (x, y)$ such that $2Q = P = (0, y_0)$. We want to find conditions for the coordinates of one such Q (and hence all four) to be rational. Now a point Q on the elliptic curve satisfies $2Q = P$ if and only if the tangent line to the curve at Q passes through $-P = (0, -y_0)$. That is, the four possible points Q are obtained geometrically by drawing the four distinct lines emanating from $-P$ which are tangent to the curve.

We readily verify that the coordinates (x, y) are rational if and only if the slope of the line from $-P$ to Q is rational. The "only if" is immediate. Conversely, if this slope m is rational, then the x-coordinate of Q, which is the double root of the cubic $(mx - y_0)^2 = (x - e_1)(x - e_2)(x - e_3)$, must also be rational. (Explicitly, $x = (e_1 + e_2 + e_3 + m^2)/2$.) In this case the y-coordinate of Q is also rational: $y = mx - y_0$. Thus, we want to know when one (and hence all four) slopes of lines from $-P$ which are tangent to E are rational.

A number $m \in \mathbb{C}$ is the slope of a line from $-P$ which is tangent to E if and only if the following equation has a double root:

$$(mx - y_0)^2 = (x - e_1)(x - e_2)(x - e_3) = x^3 + ax^2 + bx + c, \quad (9.1)$$

with

$$a = -e_1 - e_2 - e_3, \qquad b = e_1e_2 + e_1e_3 + e_2e_3, \qquad c = -e_1e_2e_3 = y_0^2, \quad (9.2)$$

where the last equality $c = y_0^2$ comes from the fact that $(0, y_0)$ is on the curve

$y^2 = x^3 + ax^2 + bx + c$. Now if we simplify (9.1) and factor out x, our condition becomes: the following quadratic equation has a double root:

$$x^2 + (a - m^2)x + (b + 2my_0) = 0.$$

This is equivalent to saying that its discriminant must vanish, i.e.,

$$(a - m^2)^2 - 4(b + 2my_0) = 0. \tag{9.3}$$

Thus, our task is to determine when one (and hence all four) roots of this quartic polynomial in m are rational.

We want to find a condition in terms of the e_i's (namely, our claim is that an equivalent condition is: $-e_i \in \mathbb{Q}^2$). In (9.3), the a and b are symmetric polynomials in the e_i, but the y_0 is not. However, y_0 is a symmetric polynomial in the $\sqrt{e_i}$. That is, we introduce f_i satisfying $f_i^2 = -e_i$. There are two possible choices for f_i, unless $e_i = 0$. Choose the f_i in any of the possible ways, subject to the condition that $y_0 = f_1 f_2 f_3$. If all of the e_i are nonzero, this means that the sign of f_1 and f_2 are arbitrary, and then the sign of f_3 is chosen so that y_0 and $f_1 f_2 f_3$ are the same square root of $-e_1 e_2 e_3$. If, say, $e_3 = 0$, then either choice can be made for the sign of f_1, f_2, and of course $f_3 = 0$. In all cases there are four possible choices of the f_i's consistent with the requirement that $y_0 = f_1 f_2 f_3$. Once we fix one such choice f_1, f_2, f_3, we can list the four choices as follows (here we're supposing that e_1 and e_2 are nonzero):

$$f_1, f_2, f_3; \qquad f_1, -f_2, -f_3; \qquad -f_1, f_2, -f_3; \qquad -f_1, -f_2, f_3. \tag{9.4}$$

The advantage of going from the e_i's to the f_i's is that now the coefficients of our equation (9.3) are symmetric functions of f_1, f_2, f_3. More precisely, if we set $s_1 = f_1 + f_2 + f_3$, $s_2 = f_1 f_2 + f_1 f_3 + f_2 f_3$, $s_3 = f_1 f_2 f_3$, the elementary symmetric functions, then

$$\begin{aligned} a &= f_1^2 + f_2^2 + f_3^2 = s_1^2 - 2s_2; \\ b &= f_1^2 f_2^2 + f_1^2 f_3^2 + f_2^2 f_3^2 = s_2^2 - 2s_1 s_3; \\ y_0 &= s_3. \end{aligned}$$

Thus, equation (9.3) becomes

$$\begin{aligned} 0 &= (m^2 - s_1^2 + 2s_2)^2 - 4(s_2^2 - 2s_1 s_3 + 2ms_3) \\ &= (m^2 - s_1^2)^2 + 4s_2(m^2 - s_1^2) - 8s_3(m - s_1). \end{aligned} \tag{9.5}$$

We see at a glance that the polynomial in (9.5) is divisible by $m - s_1$, i.e., $m = s_1 = f_1 + f_2 + f_3$ is a root. Since we could have made three other choices for the signs of the f_i, the other roots must correspond to these choices, i.e., the four solutions of equation (9.3) are:

$$\begin{aligned} m_1 &= f_1 + f_2 + f_3, \qquad & m_2 &= f_1 - f_2 - f_3, \\ m_3 &= -f_1 + f_2 - f_3, \qquad & m_4 &= -f_1 - f_2 + f_3. \end{aligned} \tag{9.6}$$

We want to know whether the four values in (9.6) are rational. Clearly, if all of the f_i are rational, then so are the m_i. Conversely, suppose the m_i are rational. Then $f_1 = (m_1 + m_2)/2$, $f_2 = (m_1 + m_3)/2$, and $f_3 = (m_1 + m_4)/2$ are rational. The conclusion of this string of equivalent conditions is: the coordinates (x, y) of a point Q for which $2Q = P$ are rational if and only if the $f_i = \sqrt{-e_i}$ are rational. This proves Proposition 20. □

Finally, we note that Proposition 20 holds with $\mathbb{Q}$ replaced by any field K not of characteristic 2. Essentially the same proof applies. (We need only take care to use algebraic rather than geometric arguments, for example, when reducing to the case $P = (0, y_0)$.)

PROBLEMS

1. Prove that for f odd, any $\mathbb{F}_{p^f}$-point of order 3 on the elliptic curve E_n: $y^2 = x^3 - n^2x$ is actually an $\mathbb{F}_p$-point; prove that there are at most three such points if $p \equiv 3 \pmod 4$; and find a fairly good sufficient condition on p and f which ensures nine $\mathbb{F}_{p^f}$-points of order 3.

2. For each of the following values of q, find the order and type of the group of $\mathbb{F}_q$-points on the elliptic curve $E_1 : y^2 = x^3 - x$. In all cases, find the type directly, if necessary checking how many points have order 3 or 4. Don't "peek" at the later problems.
   (a) All odd primes from 3 to 23.
   (b) 9
   (c) 27
   (d) 71
   (e) 11^3.

3. Find the type of the group of $\mathbb{F}_p$-points on the elliptic curve E_5: $y^2 = x^3 - 25x$ for all odd primes p of good reduction up to 23.

4. Prove that for nonzero $a \in \mathbb{Q}$ the equation $y^2 = x^3 - a$ determines an elliptic curve over any field K whose characteristic p does not divide 6 or the numerator or denominator of a; and that it has $q + 1$ $\mathbb{F}_q$-points if $q \equiv 2 \pmod 3$.

5. Prove that there are exactly 3 $\mathbb{F}_q$-points of order 3 on the elliptic curve in Problem 4 if $q \equiv 2 \pmod 3$.

6. For all odd primes p from 5 to 23, find the order and type of the group of $\mathbb{F}_p$-points on the elliptic curve $y^2 = x^3 - 1$.

7. Prove that the torsion subgroup of the group of $\mathbb{Q}$-points on the elliptic curve $y^2 = x^3 - a$ has order dividing 6 and that its order is equal to:
   (a) 6 if $a = -b^6$ for some $b \in \mathbb{Q}$;
   (b) 2 if $a = c^3$ for some $c \in \mathbb{Q}$ with c not of the form $-b^2$;
   (c) 3 if either $a = -d^2$ for some $d \in \mathbb{Q}$ with d not of the form b^3, or if $a = 432b^6$ for some $b \in \mathbb{Q}$;
   (d) 1 otherwise.

8. Show that the correspondence constructed in Problem 2 of §I.2 gives a one-to-one correspondence between right triangles as in Proposition 19 and pairs $\pm P$ of

non-identity elements of the quotient group $E_n(\mathbb{Q})/E_n(\mathbb{Q})_{\text{torsion}}$, which is isomorphic to $2E_n(\mathbb{Q})$ under the map $P \mapsto 2P$. See Problem 2(b) of §I.7.

In the problems below, we illustrate how more information can be obtained using two additional tools: (1) the complex multiplication automorphism $(x, y) \mapsto (-x, \sqrt{-1}y)$ of the group of K-points of the elliptic curve $y^2 = x^3 - n^2x$ if K contains a square root of -1; (2) the action of $\text{Gal}(K^{\text{alg cl}}/K)$ on the coordinates of the $K^{\text{alg cl}}$-points.

9. Suppose that $q \equiv 3 \pmod 4$, and l is an odd prime. Prove that:
   (a) there are at most l $\mathbb{F}_q$-points of order l on the elliptic curve $y^2 = x^3 - n^2x$, and there are at most eight $\mathbb{F}_q$-points of order 4;
   (b) the group of $\mathbb{F}_q$-points is the product of a group of order 2 and a cyclic group of order $(q + 1)/2$.

10. Suppose that $q \equiv 2 \pmod 3$, $2 \nmid N$, $3 \nmid N$. Prove that there are at most N $\mathbb{F}_q$-points of order N on the elliptic curve $y^2 = x^3 - a$.

11. Suppose that $q \equiv 1 \pmod 4$, and $l \equiv 3 \pmod 4$ is a prime not equal to p. Let (l^α, l^β) be the l-part of the type of the group of $\mathbb{F}_q$-points on the elliptic curve $y^2 = x^3 - n^2x$. Prove that $\alpha = \beta$. If $l = 2$, prove that $\alpha = \beta$ or $\alpha = \beta \pm 1$.

12. The group of K-points on an elliptic curve is analogous to the multiplicative group K^*. In Problem 11 of §I.7, we saw that for $K = \mathbb{C}$, as $a \to 0$ the elliptic curve $y^2 = (x^2 - a)(x + 1)$ "becomes" the multiplicative group $\mathbb{C}^*$. Now let K be the finite field $\mathbb{F}_q$. In this problem we work with K^*, and in the next problem we work with the group of K-points on an elliptic curve. Let l be a prime not equal to p, and suppose that $\mathbb{F}_q$ contains all l-th roots of 1, i.e., $q = p^f \equiv 1 \pmod l$.
    (a) Show that the splitting field of $x^l - a$, where $a \in \mathbb{F}_q$, has degree either 1 or l over $\mathbb{F}_q$.
    (b) Show that the subfield of $\mathbb{F}_q^{\text{alg cl}}$ generated by all l^{M+1}-th roots of 1 is $\mathbb{F}_{q^{l^{M'}}}$, where $M' \leq M$.
    (c) (For readers who know about l-adic numbers.) Construct an isomorphism between the additive group $\mathbb{Z}_l$ of l-adic integers and the galois group over $\mathbb{F}_q$ of the field extension generated by all l-th power division points (i.e., l-th power roots of unity).

13. Now let E be an elliptic curve defined over $\mathbb{F}_q$. Suppose that there are l^2 $\mathbb{F}_q$-points of order l.
    (a) Let A be an $\mathbb{F}_q$-point, and let $\mathbb{F}_{q^r}$ be the extension of $\mathbb{F}_q$ generated by the coordinates of a solution α to the equation $l\alpha = A$ (i.e., $\mathbb{F}_{q^r}$ is the smallest extension of $\mathbb{F}_q$ containing such an α). Show that there are l^2 $\mathbb{F}_{q^r}$-points α_i such that $l\alpha_i = A$.
    (b) Fix an $\mathbb{F}_{q^r}$-point α such that $l\alpha = A$. Prove that the map $\sigma \mapsto \sigma(\alpha) - \alpha$ gives an imbedding of $\text{Gal}(\mathbb{F}_{q^r}/\mathbb{F}_q)$ into the group of points of order l on E.
    (c) Show that $r = 1$ or l.
    (d) What is the field extension of $\mathbb{F}_q$ generated by all points of order l^M, $M = 1, 2, \ldots$? What is its galois group?

CHAPTER II

The Hasse–Weil L-Function of an Elliptic Curve

At the end of the last chapter, we used reduction modulo p to find some useful information about the elliptic curves $E_n: y^2 = x^3 - n^2x$ and the congruent number problem. We considered E_n as a curve over the prime field $\mathbb{F}_p$ where $p \nmid 2n$; used the easily proved equality $\#E_n(\mathbb{F}_p) = p + 1$ when $p \equiv 3 \pmod 4$; and, by making use of infinitely many such p, were able to conclude that the only rational points of finite order on E_n are the four obvious points of order two. This then reduced the congruent number problem to the determination of whether r, the rank of $E_n(\mathbb{Q})$, is zero or greater than zero.

Determining r is much more difficult than finding the torsion group. Some progress can be made using the number of $\mathbb{F}_q$-points. But the progress does not come cheaply. First of all, we will derive a formula for $\#E_n(\mathbb{F}_q)$ for any prime power $q = p^r$. Next, we will combine these numbers $N_r = N_{r,p} = \#E_n(\mathbb{F}_{p^r})$ into a function which is analogous to the Riemann zeta-function (but more complicated). The behavior of this complex-analytic function near the point 1 is intimately related to the group of rational points.

Before introducing this complex-analytic function, which is defined using all of the $N_{r,p}$, we introduce a much simpler function, called the "congruence zeta-function", which is built up from the $N_r = N_{r,p}$ for a fixed prime p.

§1. The congruence zeta-function

Given any sequence N_r, $r = 1, 2, 3, \ldots$, we define the corresponding "zeta-function" by the formal power series

$$Z(T) \underset{\text{def}}{=} \exp\left(\sum_{r=1}^{\infty} N_r \frac{T^r}{r}\right), \qquad \text{where} \quad \exp(u) \underset{\text{def}}{=} \sum_{k=0}^{\infty} \frac{u^k}{k!}. \tag{1.1}$$

At first glance, it might seem simpler to define $Z(T)$ as $\Sigma N_r T^r$; however, the above definition has crucial properties which make it the most useful one (see the problems below).

Let K be a field. Let $\mathbb{A}_K^m$ denote the set of m-tuples of elements of K. By an "affine algebraic variety in m-dimensional space over K" we mean a system of polynomial equations of the form $f_j(x_1, \ldots, x_m) = 0$, where $f_j \in K[x_1, \ldots, x_m]$. For example, a conic section is a system of two equations

$$f_1(x, y, z) = x^2 + y^2 - z^2 = 0; \qquad f_2(x, y, z) = ax + by + cz + d = 0$$

in 3-dimensional space over $\mathbb{R}$. If L is any field extension of K, the "L-points" of the variety are the m-tuples $(x_1, \ldots, x_m) \in \mathbb{A}_L^m$ for which all of the polynomials f_j vanish.

By a "projective variety in m-dimensional space over K" we mean a system of *homogeneous* polynomial equations $f_j(x_0, x_1, \ldots, x_m)$ in $m + 1$ variables. If L is a field extension of K, the "L-points" of the projective variety are the points in $\mathbb{P}_L^m$ (i.e., equivalence classes of $m + 1$-tuples $(x_0, \ldots, x_m)$, where $(x_0, \ldots, x_m) \sim (\lambda x_0, \ldots, \lambda x_m)$, $\lambda \in L^*$) at which all of the f_j vanish. For example, in the last chapter we studied the $\mathbb{F}_q$-points of the elliptic curve defined in $\mathbb{P}_{\mathbb{F}_p}^2$ by the single equation $f(x, y, z) = y^2 z - x^3 + n^2 x z^2 = 0$. (Note: Here $x_0 = z$, $x_1 = x$, $x_2 = y$ are variables for a projective variety in $\mathbb{P}_K^2$, while in the last paragraph $x_1 = x$, $x_2 = y$, $x_3 = z$ were variables for an affine variety in $\mathbb{A}_K^3$.)

If we have a projective variety, by setting $x_0 = 1$ in the f_j we obtain an affine variety whose L-points correspond to the $m + 1$-tuples with nonzero first coordinate. The remaining L-points of the projective variety will be the projective variety in $\mathbb{P}_K^{m-1}$ obtained by setting $x_0 = 0$ in all of the equations and considering the equivalence classes of m-tuples $(x_1, \ldots, x_m)$ which satisfy the resulting equations. For example, the elliptic curve with equation $y^2 z - x^3 + n^2 x z^2$ consists of the affine points—the solutions of $y^2 = x^3 - n^2 x$—and the points (x, y) of $\mathbb{P}_K^1$ for which $-x^3 = 0$, i.e., the single point $(0, 1)$ on the line at infinity $z = 0$.

Let V be an affine or projective variety defined over $\mathbb{F}_q$. For any field $K \supset \mathbb{F}_q$, we let $V(K)$ denote the set of K-points of V. By the "congruence zeta-function of V over $\mathbb{F}_q$" we mean the zeta-function corresponding to the sequence $N_r = \# V(\mathbb{F}_{q^r})$. That is, we define

$$Z(V/\mathbb{F}_q; T) \underset{\text{def}}{=} \exp\left(\sum_{r=1}^{\infty} \# V(\mathbb{F}_{q^r}) T^r / r\right). \tag{1.2}$$

Of course, N_r is finite, in fact, less than the total number of points in $\mathbb{A}_{\mathbb{F}_{q^r}}^m$ (in the affine case) or $\mathbb{P}_{\mathbb{F}_{q^r}}^m$ (in the projective case).

We shall be especially interested in the situation when V is an elliptic curve defined over $\mathbb{F}_q$. This is a special case of a smooth projective plane curve. A projective plane curve defined over a field K is a projective variety given in $\mathbb{P}_K^2$ by one homogeneous equation $f(x, y, z) = 0$. Such a curve is said to be "smooth" if there is no $K^{\text{alg cl}}$-point at which all partial derivatives

vanish. This agrees with the usual definition when $K = \mathbb{C}$ ("has a tangent line at every point").

It turns out that the congruence zeta-function of *any* elliptic curve E defined over $\mathbb{F}_q$ has the form

$$Z(E/\mathbb{F}_q; T) = \frac{1 - 2a_E T + qT^2}{(1 - T)(1 - qT)}, \tag{1.3}$$

where only the integer $2a_E$ depends on E. We shall soon prove this in the case of the elliptic curve $E_n\colon y^2 = x^3 - n^2x$. Let α be a reciprocal root of the numerator; then $1 - 2a_E T + qT^2 = (1 - \alpha T)(1 - \frac{q}{\alpha}T)$. If one takes the logarithmic derivative of both sides of (1.3) and uses the definition (1.1), one easily finds (see problems below) that the equality (1.3) is equivalent to the following formula for $N_r = \#E(\mathbb{F}_{q^r})$:

$$N_r = q^r + 1 - \alpha^r - (q/\alpha)^r. \tag{1.4}$$

As a special case of (1.4) we have

$$N_1 = \#E(\mathbb{F}_q) = q + 1 - \alpha - \frac{q}{\alpha} = q + 1 - 2a_E. \tag{1.5}$$

Thus, if we know that $Z(E/\mathbb{F}_q; T)$ must have the form (1.3), then we can determine a_E merely by counting the number of $\mathbb{F}_q$-points. This will give us $Z(E/\mathbb{F}_q; T)$, the value of α, and all of the values $N_r = \#E(\mathbb{F}_{q^r})$ by (1.4). In other words, in the case of an elliptic curve, the number of $\mathbb{F}_q$-points determines the number of $\mathbb{F}_{q^r}$-points for all r. This is an important property of elliptic curves defined over finite fields. We shall prove it in the special case $y^2 = x^3 - n^2x$.

It will also turn out that α is a quadratic imaginary algebraic integer whose complex absolute value is $\sqrt{q}$. In the case $y^2 = x^3 - n^2x$, it will turn out that α is a square root of $-q$ if $q \equiv 3 \pmod 4$, and is of the form $a + bi$, $a, b \in \mathbb{Z}$, $a^2 + b^2 = q$, if $q \equiv 1 \pmod 4$.

This situation is a special case of a much more general fact concerning smooth projective algebraic varieties over finite fields. The general result was conjectured by André Weil in [Weil 1949], and the last and most difficult part was proved by Pierre Deligne in 1973. (For a survey of Deligne's proof, see [Katz 1976a].) We shall not discuss it, except to state what it says in the case of a smooth projective *curve* (one-dimensional variety):

(i) $Z(V/\mathbb{F}_q; T)$ is a rational function of T (this is true for any variety without the smoothness assumption) which for a smooth curve has the form $P(T)/(1 - T)(1 - qT)$. Here $P(T)$ has coefficients in $\mathbb{Z}$ and constant term 1 (equivalently, its reciprocal roots are algebraic integers).

(ii) If V was obtained by reducing modulo p a variety $\tilde{V}$ defined over $\mathbb{Q}$, then $\deg P = 2g$ is twice the genus ("Betti number") of the complex analytic manifold $\tilde{V}$. Intuitively, g is the "number of handles" in the corresponding Riemann surface. An elliptic curve has $g = 1$, and the Riemann surface in Fig. II.1 has $g = 3$.

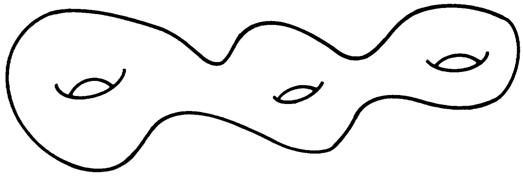


Figure II.1

(iii) If α is a reciprocal root of the numerator, then so is q/α.
(iv) All reciprocal roots of the numerator have complex absolute value $\sqrt{q}$.

One reason for the elegance of the Weil conjectures is the intriguing indirect connection between the "physical" properties of a curve (e.g., its number of handles as a Riemann surface when considered over $\mathbb{C}$) and the number theoretic properties (its number of points when considered over $\mathbb{F}_{q^r}$). Roughly speaking, it says that the more complicated the curve is (the higher its genus), the more N_r's you need to know before the remaining ones can be determined. In the simplest interesting case, that of elliptic curves, where $g = 1$, all of the N_r's are determined once you know N_1.

Problems

1. Show that if $N_r = N_r^* + N_r^{**}$ and $Z(T)$, $Z^*(T)$, $Z^{**}(T)$ are the corresponding zeta-functions, then $Z(T) = Z^*(T) \cdot Z^{**}(T)$; and if $N_r = N_r^* - N_r^{**}$, then $Z(T) = Z^*(T)/Z^{**}(T)$.

2. Show that if there exists a fixed set $\alpha_1, \ldots, \alpha_s, \beta_1, \ldots, \beta_t$ such that for all r we have $N_r = \beta_1^r + \cdots + \beta_t^r - \alpha_1^r - \cdots - \alpha_s^r$, then
$$Z(T) = \frac{(1 - \alpha_1 T)(1 - \alpha_2 T)\cdots(1 - \alpha_s T)}{(1 - \beta_1 T)(1 - \beta_2 T)\cdots(1 - \beta_t T)}.$$

3. Prove that if $|N_r| < CA^r$ for some constants C and A, then the power series $Z(T)$ converges in the open disc of radius $1/A$ in the complex plane.

4. Show that if $N_r = \begin{cases} 1, & r \text{ even;} \\ 0, & r \text{ odd,} \end{cases}$ then $Z(T)$ is *not* a rational function; but if $N_r = \begin{cases} 2, & r \text{ even;} \\ 0, & r \text{ odd,} \end{cases}$ then $Z(T)$ is rational. In the latter case, interpret N_r as the number of $\mathbb{F}_{p^r}$-solutions of some equation.

5. The Bernoulli polynomials $B_r(x) \in \mathbb{Q}[x]$ have the properties: (i) deg $B_r = r$; (ii) for all M, $B_r(M) - B_r(0) = r(1^{r-1} + 2^{r-1} + \cdots + (M-1)^{r-1})$. Now for fixed M let $N_{r-1} = \frac{1}{r}(B_r(M) - B_r(0))$. Find the corresponding $Z(T)$. (Cultural note: $B_1(x) = x - \frac{1}{2}$, $B_2(x) = x^2 - x + \frac{1}{6}$, etc.; they are uniquely determined by properties (i) and (ii) along with the normalization requirement that $\int_0^1 B_r(x)dx = 0$ for $r \geq 1$. One way to define them is by equating terms in the relation: $te^{tx}/(e^t - 1) = \sum_{r=0}^{\infty} B_r(x)t^r/r!$.)

6. Suppose that l is a prime, q is a power of another prime p, $q \equiv 1 \pmod{l}$, $q \not\equiv 1 \pmod{l^2}$.
   (a) For fixed M, let $N_r = \#\{x \in \mathbb{F}_{q^r} | x^{l^M} = 1\}$. Find the corresponding $Z(T)$.
   (b) Now let $N_r = \#\{x \in \mathbb{F}_{q^r} | x^{l^M} = 1 \text{ for some } M\}$. Find the corresponding zeta-function. Is it rational?

7. A special case of an affine or projective variety V is the entire space, corresponding to the empty set of equations. Let $\mathbb{A}_K^m$ denote m-dimensional affine space (the usual space of m-tuples of numbers in the field K), and let $\mathbb{P}_K^m$ denote projective space, as usual.
   (a) What is $Z(\mathbb{A}_{\mathbb{F}_q}^m/\mathbb{F}_q; T)$?
   (b) Find $Z(\mathbb{P}_{\mathbb{F}_q}^m/\mathbb{F}_q; T)$ by writing $\mathbb{P}_K^m$ as a disjoint union of $\mathbb{A}_K^k$, $k = m, m-1, \ldots, 0$, and using Problem 1.
   (c) Also find $Z(\mathbb{P}_{\mathbb{F}_q}^m/\mathbb{F}_q; T)$ by counting equivalence classes of $(m+1)$-tuples, and check that your answers agree.

8. Show that, if V is a variety in $\mathbb{A}_{\mathbb{F}_q}^m$ or $\mathbb{P}_{\mathbb{F}_q}^m$, then $Z(V/\mathbb{F}_q; T)$ converges for $|T| < q^{-m}$.

9. If one wants to prove that $Z(V/\mathbb{F}_q; T) \in \mathbb{Z}[[T]]$ with constant term 1 for any affine or projective variety V, show that it suffices to prove this when V is any *affine* variety. Then show that it suffices to prove this when V is given by a single equation. Show that the rationality assertion $Z(V/\mathbb{F}_q; T) \in \mathbb{Q}(T)$ can also be reduced to the case of an affine variety V defined by a single equation. A variety defined by a single equation is called a "hypersurface".

10. Find the zeta-function of the curve $y^2 = x^3 - n^2x$ in $\mathbb{P}_{\mathbb{F}_q}^2$ if $p|2n$, i.e., p is *not* a prime of good reduction.

11. Find the zeta-function of the hypersurface in $\mathbb{A}_{\mathbb{F}_q}^4$ defined by $x_1x_2 - x_3x_4 = 0$.

12. Let N_r be the number of lines in $\mathbb{P}_{\mathbb{F}_{q^r}}^3$. Find its zeta-function. (It is possible to view the set of k-dimensional subspaces in $\mathbb{P}_K^m$ as a variety, called the grassmannian; in our case $k = 1$, $m = 3$.)

13. Using the form (1.3) for the zeta-function of an elliptic curve, where the numerator has reciprocal root α, show that N_r is equal to the norm of $1 - \alpha^r$. Now, in the situation of Problem 13 of §I.9, suppose that E has l^2 $\mathbb{F}_q$-points of order l, and no $\mathbb{F}_q$-points of exact order l^2. Prove that the field extension of $\mathbb{F}_q$ generated by the coordinates of the points of order l^{M+1} is $\mathbb{F}_{q^{l^M}}$. (Note the close analogy with the multiplicative group $\mathbb{F}_q^*$, with $q \equiv 1 \pmod{l}$ but $q \not\equiv 1 \pmod{l^2}$, where the field generated by all l^{M+1}-th roots of unity is $\mathbb{F}_{q^{l^M}}$.)

14. Let V be an affine algebraic variety defined over K by equations $f_j(x_1, \ldots, x_m) = 0$. By the coordinate ring $R(V)$ we mean the quotient ring of $K[x_1, \ldots, x_m]$ by the ideal generated by all of the f_j. Let $P = (a_1, \ldots, a_m)$ be a $K^{\text{alg cl}}$-point on V. Let $L = K(a_1, \ldots, a_m)$ be the finite extension of K generated by the coordinates of P. L is called the residue field of P, and its degree over K is called the residue degree.
   (a) Show that the map $x_i \mapsto a_i$ is well-defined on $R(V)$, and extends to a homomorphism whose kernel is a maximal ideal $m(P)$ in $R(V)$. (It is not hard to prove that every maximal ideal of $R(V)$ arises in this way.)
   (b) Show that $m(P') = m(P)$ if and only if there is an isomorphism from L to L'

(the residue fields of P and P', respectively) which takes a_i to a'_i. Thus, the maximal ideal $m(P)$ corresponds to d different $K^{\text{alg cl}}$-points P on V, where $d = [R(V)/m(P): K]$ is the residue degree of any of the points P.

15. In the situation of Problem 14, let $K = \mathbb{F}_q$. For a given $K^{\text{alg cl}}$-point P, the residue field is $\mathbb{F}_{q^d}$ for some d. Then P contributes 1 to each N_r for which r is a multiple of d. That is, the contribution of P to the exponent in the definition of the zeta-function is $\Sigma_{k=1}^{\infty} T^{kd}/kd$. Then $Z(V/\mathbb{F}_q; T)$ is exp of the sum of all contributions from the different $K^{\text{alg cl}}$-points P. Group together all points corresponding to a given maximal ideal, and express $Z(V/\mathbb{F}_q; T)$ as the product over all maximal ideals $\mathbf{m}$ of $(1 - T^{\deg \mathbf{m}})^{-1}$. Then show that the zeta-function belongs to $1 + T\mathbb{Z}[[T]]$. (Cultural note: If we make the change of variables $T = q^{-s}$, and define $\text{Norm}(\mathbf{m})$ to be the number of elements in the residue field, i.e., $\text{Norm}(\mathbf{m}) = q^{\deg \mathbf{m}}$, then we have $Z(V/\mathbb{F}_q; q^{-s}) = \Pi_{\mathbf{m}}(1 - \text{Norm}(\mathbf{m})^{-s})^{-1}$, which is closely analogous to the Euler product for the Dedekind zeta-function of a number field: $\zeta_K(s) = \Pi_{\mathfrak{p}}(1 - \text{Norm}(\mathfrak{p})^{-s})^{-1}$, in which the product is over all nonzero prime ideals of the ring of integers in the field K. In a number ring, a nonzero prime ideal is the same as a maximal ideal.)

16. Prove that if $Z(V/\mathbb{F}_q; T) \in \mathbb{Q}(T)$, then the numerator and denominator are in $1 + T\mathbb{Z}[T]$ (equivalently, the α's and β's in Problem 2 are algebraic integers).

§2. The zeta-function of E_n

We now return to our elliptic curve E_n, which is the curve $y^2 = x^3 - n^2x$, where n is a squarefree positive integer. More precisely, E_n is the projective completion of this curve, i.e., we also include the point at infinity. E_n is an elliptic curve over any field K whose characteristic does not divide $2n$, and, as we have seen, it is sometimes useful to take $K = \mathbb{F}_p$, or more generally $K = \mathbb{F}_q$. The purpose of this section is to express the number of $\mathbb{F}_q$-points on E_n in terms of "Jacobi sums".

To do this, we first transform the equation of E_n to a "diagonal form". We say that a hypersurface $f(x_1, \ldots, x_n) = 0$ in $\mathbb{A}_K^m$ is "diagonal" if each monomial in f involves at most one of the variables, and each variable occurs in at most one monomial. For example, the "Fermat curve" $x^d + y^d = 1$ is diagonal. It turns out that diagonal hypersurfaces lend themselves to easy computation of the N_r (much in the same way that multiple integrals are much easier to evaluate when the variables separate). We shall not treat the general case, but only the one we need to evaluate $N_r = \#E_n(\mathbb{F}_{q^r})$. (For a general treatment of diagonal hypersurfaces, see [Weil 1949] or [Ireland and Rosen 1990, Chapter 11].)

We first show a relation between points on E_n: $y^2 = x^3 - n^2x$ and points on the curve E'_n: $u^2 = v^4 + 4n^2$. As usual, we suppose that $p \nmid 2n$. First suppose that (u, v) is on E'_n. Then it is easy to check that the point $(x, y) = (\frac{1}{2}(u + v^2), \frac{1}{2}v(u + v^2))$ is on E_n. Conversely, if (x, y) is on E_n and its x-coordinate is nonzero, then we check that the point $(u, v) = (2x - y^2/x^2, y/x)$ is on E'_n.

Moreover, these two maps are inverse to one another. In other words, we have a one-to-one correspondence between points on E_n' and points on $E_n - \{(0, 0)\}$. Let N' be the number of $\mathbb{F}_q$-solutions (u, v) to $u^2 = v^4 + 4n^2$. Then the points on our elliptic curve consist of $(0, 0)$, the point at infinity, and the N' points corresponding to the pairs (u, v). In other words, $N_1 = \#E_n(\mathbb{F}_q)$ is equal to $N' + 2$. So it remains to compute N'. The advantage of the equation $u^2 = v^4 + 4n^2$ is that it is diagonal.

The basic ingredients in determining the number of points on a diagonal hypersurface are the Gauss and Jacobi sums over finite fields. We shall now define them and give their elementary properties.

Let $\psi\colon \mathbb{F}_q \to \mathbb{C}^*$ be a nontrivial additive character, i.e., a nontrivial homomorphism from the additive group of the finite field to the multiplicative group of complex numbers. (Since $\mathbb{F}_q$ is finite, the image must consist of roots of unity.) In what follows, we shall always define $\psi(x) = \xi^{\operatorname{Tr} x}$, where $\xi = e^{2\pi i/p}$, and Tr is the trace from $\mathbb{F}_q$ to $\mathbb{F}_p$. Since the trace is a nontrivial additive map, and its image is $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$, we obtain in this way a nontrivial additive character.

Now let $\chi\colon \mathbb{F}_q^* \to \mathbb{C}^*$ be any multiplicative character, i.e., a group homomorphism from the multiplicative group of the finite field to the multiplicative group of nonzero complex numbers. In what follows, the additive character ψ will be fixed, as defined above, but χ can vary.

We define the Gauss sum (depending on the variable χ) by the formula

$$g(\chi) = \sum_{x \in \mathbb{F}_q} \chi(x)\psi(x)$$

(where we agree to take $\chi(0) = 0$ for *all* χ, even the trivial multiplicative character). We define the Jacobi sum (depending on two variable multiplicative characters) by the formula

$$J(\chi_1, \chi_2) = \sum_{x \in \mathbb{F}_q} \chi_1(x)\chi_2(1 - x).$$

The proofs of the following elementary properties of Gauss and Jacobi sums are straightforward, and will be left as exercises. (Here χ_{triv} denotes the trivial character, which takes all nonzero elements of $\mathbb{F}_q$ to 1; χ, χ_1, and χ_2 denote nontrivial characters; and $\bar{\chi}$ denotes the complex conjugate (also called "inverse") character of χ, whose value at x is the complex conjugate of $\chi(x)$.)

(1) $g(\chi_{\text{triv}}) = -1;\quad J(\chi_{\text{triv}}, \chi_{\text{triv}}) = q - 2;\quad J(\chi_{\text{triv}}, \chi) = -1;$
$J(\chi, \bar{\chi}) = -\chi(-1);\quad J(\chi_1, \chi_2) = J(\chi_2, \chi_1);$
(2) $g(\chi)\cdot g(\bar{\chi}) = \chi(-1)q;\quad |g(\chi)| = \sqrt{q};$
(3) $J(\chi_1, \chi_2) = g(\chi_1)g(\chi_2)/g(\chi_1\chi_2)\quad \text{if}\quad \chi_2 \neq \bar{\chi}_1.$

We now proceed to the computation of the number N' of $u, v \in \mathbb{F}_q$ satisfying $u^2 = v^4 + 4n^2$. The key observation in computing N' is that for any $a \neq 0$ in $\mathbb{F}_q$ and any m dividing $q - 1$, the number of solutions $x \in \mathbb{F}_q$ to the equation $x^m = a$ is given by:

$$\#\{x^m = a\} = \sum_{\chi^m=1} \chi(a), \tag{2.1}$$

where the sum is over all multiplicative characters whose m-th power is the trivial character. Namely, both sides of (2.1) equal m if a is an m-th power in $\mathbb{F}_q$ and equal 0 otherwise; the detailed proof will be left as a problem below.

By Proposition 16 of the last chapter, we know that $N_1 = q + 1$ if $q \equiv 3 \pmod 4$. In what follows, we shall suppose that $q \equiv 1 \pmod 4$.

In counting the pairs (u, v), we count separately the pairs where either u or v is zero. Thus, we write

$$\begin{aligned} N' = {} & \#\{u \in \mathbb{F}_q | u^2 = 4n^2\} + \#\{v \in \mathbb{F}_q | 0 = v^4 + 4n^2\} \\ & + \#\{u, v \in \mathbb{F}_q^* | u^2 = v^4 + 4n^2\}. \end{aligned} \tag{2.2}$$

The first term in (2.2) is obviously 2 (recall that we are assuming that $p \nmid 2n$). We use (2.1) to evaluate the second term. Let χ_4 be one of the characters of $\mathbb{F}_q^*$ having exact order 4, i.e., $\chi_4(g) = i$ for some generator g of the cyclic group $\mathbb{F}_q^*$. Then, by (2.1), the second term in (2.2) equals

$$\sum_{j=1}^{4} \chi_4^j(-4n^2) = 2 + 2\chi_4(-4n^2) \tag{2.3}$$

(where we use the fact that $-4n^2$ is a square in $\mathbb{F}_q^*$). Finally, we evaluate the third term in (2.2). Let χ_2 denote the nontrivial character of order 2 (i.e., $\chi_2 = \chi_4^2$). Using (2.1) again, we can write the third term in (2.2) as

$$\sum_{\substack{a, b \in \mathbb{F}_q^* \\ a=b+4n^2}} \#\{u^2 = a\} \cdot \#\{v^4 = b\} = \sum_{a \in \mathbb{F}_q^*, a-4n^2 \neq 0} \ \sum_{\substack{j=1,2,3,4 \\ k=1,2}} \chi_2^k(a)\chi_4^j(a - 4n^2).$$

Note that since $\chi_4^j(0) = 0$, we can drop the condition $a - 4n^2 \neq 0$ on the right. We now make the change of variable $x = a/4n^2$ in the first summation on the right. As a result, after we reverse the order of summation, the right side becomes

$$\sum_{\substack{j=1,2,3,4 \\ k=1,2}} \chi_4^j(-4n^2) \sum_{x \in \mathbb{F}_q^*} \chi_2^k(x)\chi_4^j(1 - x) = \sum_{\substack{j=1,2,3,4 \\ k=1,2}} \chi_4^j(-4n^2) J(\chi_2^k, \chi_4^j).$$

Finally, bringing together the three terms in (2.2) and using property (1) of Jacobi sums when χ_2^k or χ_4^j is trivial or they are conjugate to one another, we obtain:

$$\begin{aligned} N' &= 4 + 2\chi_4(-4n^2) + \sum_{j=1,3} \chi_4^j(-4n^2) J(\chi_2, \chi_4^j) + q - 2 + 3 \cdot (-1) \\ &\quad + 2\chi_4(-4n^2) \cdot (-1) \\ &= q - 1 + \chi_4(-4n^2)(J(\chi_2, \chi_4) + J(\chi_2, \bar{\chi}_4)). \end{aligned} \tag{2.4}$$

In the problems we show that $\chi_4(-4) = 1$. Hence, $\chi_4(-4n^2) = \chi_2(n)$. Thus, if we set

$$\alpha = \alpha_{n,q} \underset{\text{def}}{=} -\chi_2(n) J(\chi_2, \chi_4), \tag{2.5}$$

we conclude that

$$N_1 = \#E_n(\mathbb{F}_q) = q + 1 - \alpha - \bar{\alpha}. \tag{2.6}$$

Notice that α is an algebraic integer in $\mathbb{Q}(i)$, since the values of χ_2 and χ_4 in the definition of $J(\chi_2, \chi_4)$ are all ± 1, $\pm i$. We now pin down the Gaussian integer $\alpha = a + bi$, at least in the case when $q = p$ is a prime congruent to 1 mod 4 or $q = p^2$ is the square of a prime congruent to 3 mod 4. By property (3) relating Jacobi to Gauss sums, we have

$$\alpha = -\chi_2(n)g(\chi_2)g(\chi_4)/g(\bar{\chi}_4),$$

and hence, by property (2), we have $|\alpha|^2 = a^2 + b^2 = q$. In the two cases $q = p \equiv 1 \pmod 4$ and $q = p^2, p \equiv 3 \pmod 4$, there are very few possibilities for such an α. Namely, in the former case there are eight choices of the form $\pm a \pm bi$, $\pm b \pm ai$; and in the latter case there are the four possibilities $\pm p$, $\pm pi$. The following lemma enables us to determine which it is.

Lemma 1. *Let $q \equiv 1 \pmod 4$, and let χ_2 and χ_4 be characters of $\mathbb{F}_q^*$ of exact order 2 and 4, respectively. Then $1 + J(\chi_2, \chi_4)$ is divisible by $2 + 2i$ in the ring $\mathbb{Z}[i]$.*

PROOF. We first relate $J(\chi_2, \chi_4)$ to $J(\chi_4, \chi_4)$ by expressing both in terms of Gauss sums. By property (3), we have: $J(\chi_2, \chi_4) = J(\chi_4, \chi_4)g(\chi_2)^2/g(\chi_4)g(\bar{\chi}_4)$ $= \chi_4(-1)J(\chi_4, \chi_4)$ by property (2). Next, we write

$$J(\chi_4, \chi_4) = \sum \chi_4(x)\chi_4(1 - x) = \chi_4^2(\tfrac{p+1}{2}) + 2\sum{}' \chi_4(x)\chi_4(1 - x),$$

where Σ' is a sum over $(q - 3)/2$ elements, one from each pair x, $1 - x$, with the pair $\frac{(p+1)}{2}, \frac{(p+1)}{2}$ omitted. Notice that $\chi_4(x)$ is a power of i, and so is congruent to 1 modulo $1 + i$ in $\mathbb{Z}[i]$; thus, $2\chi_4(x)\chi_4(1 - x) \equiv 2 \pmod{2 + 2i}$. As a result, working modulo $2 + 2i$, we have $J(\chi_4, \chi_4) \equiv q - 3 + \chi_4^2(\frac{(p+1)}{2}) \equiv 2 + \chi_4(4)$ (since $q \equiv 1 \pmod 4$). Returning to $J(\chi_2, \chi_4)$, we obtain:

$$1 + J(\chi_2, \chi_4) = 1 + \chi_4(-1)J(\chi_4, \chi_4) \equiv 1 + \chi_4(-4) + 2\chi_4(-1)$$
$$\pmod{2 + 2i}.$$

Since $\chi_4(-4) = 1$, as mentioned above (and proved in the problems below), and since $2(1 + \chi_4(-1)) = 0$ or 4, it follows that $1 + J(\chi_2, \chi_4)$ is divisible by $2 + 2i$, as claimed. □

We now have the basic ingredients to prove a formula for $Z(E_n/\mathbb{F}_p; T)$.

Theorem. *Let E_n be the elliptic curve $y^2 = x^3 - n^2x$ defined over $\mathbb{F}_p$, where $p \nmid 2n$. Then*

$$Z(E_n/\mathbb{F}_p; T) = \frac{1 - 2aT + pT^2}{(1 - T)(1 - pT)} = \frac{(1 - \alpha T)(1 - \bar{\alpha}T)}{(1 - T)(1 - pT)}, \tag{2.7}$$

where $a = \operatorname{Re}\alpha$; $\alpha = i\sqrt{p}$ if $p \equiv 3 \pmod 4$; and if $p \equiv 1 \pmod 4$, then α is an element of $\mathbb{Z}[i]$ of norm p which is congruent to $(\frac{n}{p})$ modulo $2 + 2i$.

Before proving the theorem, we note that in the case $p \equiv 1 \pmod 4$ it says we choose $\alpha = a + bi$ with a odd (and b even), where the sign of a is determined by the congruence condition modulo $2 + 2i$. There are two possible choices $a + bi$ and $a - bi$; and of course the formula (2.7) does not change if we replace α by its conjugate.

PROOF. In order to obtain $Z(E_n/\mathbb{F}_p; T)$, we must let the power of p vary, and determine $N_r = \# E_n(\mathbb{F}_{p^r})$ for $p \equiv 1 \pmod 4$ and $N_{2r} = \# E_n(\mathbb{F}_{q^r})$ for $p \equiv 3 \pmod 4$, $q = p^2$ (since we know that $N_r = p^r + 1$ for odd r in that case). So we fix q equal to p in the first case and equal to p^2 in the second case (in either case $q \equiv 1 \pmod 4$), and we replace q by q^r throughout the work we did earlier to find a formula for $\# E_n(\mathbb{F}_q)$, $q \equiv 1 \pmod 4$.

Because the r is varying, we need a notation to indicate which χ_2 and χ_4 we are talking about, i.e., to indicate for which finite field they are multiplicative characters. Let $\chi_{2,1} = \chi_2$ denote the unique nontrivial character of $\mathbb{F}_q^*$ of order 2, and let $\chi_{4,1} = \chi_4$ denote a fixed character of $\mathbb{F}_q^*$ of exact order 4 (there are two, the other one being $\bar{\chi}_4$). Then by composing χ_2 or χ_4 with the norm from $\mathbb{F}_{q^r}$ to $\mathbb{F}_q$, we obtain a character of $\mathbb{F}_{q^r}^*$ of exact order 2 or 4, respectively. We denote these characters $\chi_{2,r}$ and $\chi_{4,r}$. For example, if g is a generator of $\mathbb{F}_q^*$ such that $\chi_4(g) = i$, and if g_r is a generator of $\mathbb{F}_{q^r}^*$ whose norm is g, i.e., $(g_r)^{1+q+\cdots+q^{r-1}} = g$, then we have $\chi_{4,r}(g_r) = i$. If $\mathbb{N}_r$ denotes the norm from $\mathbb{F}_{q^r}$ to $\mathbb{F}_q$, we can write our definitions:

$$\chi_{4,r} = \chi_4 \circ \mathbb{N}_r, \qquad \chi_{2,r} = \chi_2 \circ \mathbb{N}_r. \tag{2.8}$$

With these definitions, using (2.5) and (2.6), we can write:

$$\begin{gathered} \# E_n(\mathbb{F}_{q^r}) = q^r + 1 - \alpha_{n,q^r} - \bar{\alpha}_{n,q^r}, \\ \text{where} \quad \alpha_{n,q^r} = -\chi_{2,r}(n) \frac{g(\chi_{2,r}) g(\chi_{4,r})}{g(\bar{\chi}_{4,r})}. \end{gathered} \tag{2.9}$$

We now use a basic relationship, called the Hasse–Davenport relation, for Gauss sums over extensions of finite fields. The Hasse–Davenport formula is:

$$-g(\chi \circ \mathbb{N}_r) = (-g(\chi))^r. \tag{2.10}$$

The proof of this fact will be given in a series of exercises below. Applying (2.10) to the three Gauss sums in (2.9), and observing that $\chi_{2,r}(n) = \chi_2(n^r) = \chi_2(n)^r$, we conclude the following basic relationship:

$$\alpha_{n,q^r} = \alpha_{n,q}^r. \tag{2.11}$$

The theorem now follows quickly. First suppose $p \equiv 1 \pmod 4$, in which case $q = p$. Then $\chi_2(n)$ is the Legendre symbol $(\frac{n}{p})$. Using (2.5) and Lemma 1, we find that $\alpha = \alpha_{n,p}$ is a Gaussian integer of norm p which is congruent to $(\frac{n}{p})$ modulo $2 + 2i$; and, by (2.9) and (2.11),

$$N_r = p^r + 1 - \alpha^r - \bar{\alpha}^r.$$

This proves the theorem when $p \equiv 1 \pmod 4$ (see Problem 2 of §II.1).

Now suppose that $p \equiv 3 \pmod 4$, $q = p^2$. Then $\chi_2(n) = 1$, since all elements of $\mathbb{F}_p$ are squares in $\mathbb{F}_{p^2}$. Then Lemma 1 tells us that $\alpha_{n,q}$ is a Gaussian integer of norm q which is congruent to 1 mod $2 + 2i$. Of the four Gaussian integers $i^j p$, $j = 0, 1, 2, 3$, having norm q, only $\alpha_{n,q} = -p$ satisfies the congruence condition. Then, by (2.9) and (2.11), we conclude that for r even we have

$$N_r = \#E_n(\mathbb{F}_{q^{r/2}}) = p^r + 1 - (-p)^{r/2} - (-p)^{r/2}.$$

Since $N_r = p^r + 1$ for odd r, we have for any r:

$$N_r = p^r + 1 - (i\sqrt{p})^r - (-i\sqrt{p})^r.$$

This completes the proof of the theorem. □

We conclude this section by calling attention to the role Lemma 1 has played in pinning down the reciprocal roots α and $\bar{\alpha}$ in (2.7). The congruence condition in Lemma 1 will again be needed when we start working with the Hasse–Weil L-function of the elliptic curve E_n, which combines the α's for different primes p. In that context, Lemma 1 is a special case of a general fact about how Jacobi sums vary as we vary the prime p. The general case is treated in [Weil 1952].

Problems

1. Prove properties (1)–(3) of Gauss and Jacobi sums that were given in the text.

2. Let G be a finite group, and let $\check{G}$ denote the group of characters χ (i.e., of homomorphisms $\chi\colon G \to \mathbb{C}^*$). Recall that for any nontrivial $\chi \in \check{G}$, $\Sigma_{g\in G}\,\chi(g) = 0$. Notice that any fixed $g \in G$ gives a character $g\colon \chi \mapsto \chi(g)$ on the group $\check{G}$, and also on any subgroup $S \subset \check{G}$. Apply these general considerations to the case when $G = \mathbb{F}_q^*$ and S is the subgroup of characters χ such that $\chi^m = 1$. In that way prove the relation (2.1) in the text.

3. Let $q \equiv 1 \pmod 4$, and let χ_4 have exact order 4. Show that $\chi_4(4)$ and $\chi_4(-1)$ are both equal to 1 if $q \equiv 1 \pmod 8$ and equal to -1 if $q \equiv 5 \pmod 8$. Conclude that $\chi_4(-4) = 1$ in all cases.

4. Show that $g(\chi_2)^2 = (-1)^{(q-1)/2}q$. It is somewhat harder to determine which square root to take to get $g(\chi_2)$ (see [Borevich and Shafarevich 1966, pp. 349–353]). Compute $g(\chi_2)$ when $q = 3, 5, 7, 9$.

5. For $q \equiv 1 \pmod 4$, again let χ_2 be the nontrivial quadratic character, and let χ_4 and $\bar{\chi}_4$ be the two characters of exact order 4. Compute $J(\chi_2, \chi_4)$ and $J(\chi_2, \bar{\chi}_4)$ directly from the definition when $q = 5, 9, 13, 17$.

6. Show that if χ_2 is the nontrivial quadratic character of $\mathbb{F}_q^*$ and χ is any nontrivial character, then $J(\chi_2, \chi) = \chi(4)J(\chi, \chi)$.

7. Let χ_3 and $\bar{\chi}_3$ be the two characters of $\mathbb{F}_q^*$ of order 3, where $q \equiv 1 \pmod 3$. Compute $J(\chi_3, \chi_3)$ and $J(\bar{\chi}_3, \bar{\chi}_3)$ directly from the definition when $q = 7, 13$.

8. (a) Notice that we proved that the number N_r of $\mathbb{F}_{q^r}$-points on E_n is independent of n if r is even. Show this directly.
   (b) Also notice that N_r does not change if n is multiplied by an integer which is a square in $\mathbb{F}_q$. This is for the same reason that we could, without loss of generality, reduce to squarefree n when considering $\mathbb{Q}$-points. Namely, if K is any field not of characteristic 2 and if m, $n \in K^*$, construct a simple correspondence between $E_n(K)$ and $E_{nm^2}(K)$.

9. This problem concerns a more general definition of Gauss sums, examples of which will occur later in the chapter. Let R be the ring of integers in a number field K, and let I be a nonzero ideal of R. Then R/I is a finite ring. Let ψ: $R/I \to \mathbb{C}^*$ be an additive character which is nontrivial on any additive subgroup of R/I of the form J/I for any strictly larger ideal $J \supset I$ (including the "improper ideal" $J = R$, which will be the only such J if I is a prime ideal). Define the norm $\mathbb{N}I = \#(R/I)$. Let χ: $(R/I)^* \to \mathbb{C}^*$ be any multiplicative character. Take $\chi(x) = 0$ for $x \in R/I$ not prime to I. Define $g(\chi) = g(\chi, \psi) \underset{\text{def}}{=} \Sigma\, \chi(x)\psi(x)$, where the summation is over $x \in R/I$.
   (a) Prove that $\Sigma\, \chi(x)\psi(ax) = \bar{\chi}(a)g(\chi, \psi)$ for any $a \in (R/I)^*$.
   In parts (b) and (c) we suppose that χ is "primitive" modulo I. By definition, this means that, for any strictly larger ideal $J \supset I$, χ is nontrivial on the subgroup of $(R/I)^*$ consisting of elements congruent to 1 modulo J.
   (b) If χ is primitive, show that the formula in part (a) holds for all $a \in R/I$.
   (c) For χ primitive, prove that $g(\chi, \psi)g(\bar{\chi}, \psi) = \chi(-1)\mathbb{N}I$, and $|g(\chi, \psi)| = \sqrt{\mathbb{N}I}$.
   Some examples of the characters and Gauss sums in this problem are: (1) if I is a prime ideal with residue field $\mathbb{F}_q$, then property (2) of Gauss sums in the text is a special case of part (c); (2) if $R = \mathbb{Z}$ and I is the ideal (N), then χ is an ordinary Dirichlet character. $\mathbb{N}I = N$, we often take $\psi(x) = e^{2\pi i x/N}$, and "primitive" means that the value of $\chi(x)$ for $x \in (\mathbb{Z}/N\mathbb{Z})^*$ does not depend only on its residue modulo some proper divisor of N; (3) later in the chapter we will encounter examples where $R = \mathbb{Z}[i]$.

Problems 10–17 will lead to a proof of the Hasse–Davenport relation.

10. Let S be the set of all monic polynomials in $\mathbb{F}_q[x]$, and let S^{irr} denote the subset of all irreducible monic polynomials. Subscripts will indicate degree. By writing $x^{q^r} - x = \Pi_{\alpha \in \mathbb{F}_{q^r}}(x - \alpha)$, prove that $x^{q^r} - x = \Pi f$, where the product is over all f in S_d^{irr} for all d dividing r.

11. Let ψ be a nontrivial additive character and χ a multiplicative character of $\mathbb{F}_q$. If $f \in S$ is written in the form $f(x) = x^d - c_1 x^{d-1} + \cdots + (-1)^d c_d$, define a map λ: $S \to \mathbb{C}$ by $\lambda(f) = \chi(c_d)\psi(c_1)$. (If $f = 1$ is the constant function in S_0, then define $\lambda(1) = 1$.) Prove that $\lambda(f_1 f_2) = \lambda(f_1)\lambda(f_2)$ for $f_1, f_2 \in S$.

12. Prove that the Gauss sum can be written $g(\chi) = \Sigma_{f \in S_1} \lambda(f)$.

13. Suppose that $\alpha \in \mathbb{F}_{q^r}$ satisfies monic irreducible polynomial $f \in S_d^{\text{irr}}$, where $d|r$. Then show that $\lambda(f)^{r/d} = \chi_r(\alpha) \cdot \psi_r(\alpha)$, where the subscripts here indicate the characters of $\mathbb{F}_{q^r}$ obtained by composing with the norm from $\mathbb{F}_{q^r}$ to $\mathbb{F}_q$ (in the case of a multiplicative character) or with the trace from $\mathbb{F}_{q^r}$ to $\mathbb{F}_q$ (in the case of an additive character).

14. Prove that $g(\chi_r) = \Sigma_{d|r} \Sigma_{f \in S_d^{\text{irr}}} d\lambda(f)^{r/d}$.

15. Prove the power series identity $\Sigma_{f\in S}\lambda(f)T^{\deg f} = \Pi_{f\in S^{\text{irr}}}(1 - \lambda(f)T^{\deg f})^{-1}$.

16. Show that if $d > 1$, then $\Sigma_{f\in S_d}\lambda(f) = 0$.

17. Taking the logarithmic derivative of both sides in Problem 15, prove that

$$(-1)^{r-1}g(\chi)^r = \sum_{d|r}\sum_{f\in S_d^{\text{irr}}} d\lambda(f)^{r/d},$$

and conclude the proof of the Hasse–Davenport relation.

18. (a) Show that the ideal (2) in $\mathbb{Z}[i]$ is the square of the prime ideal $(1 + i)$; and that any element $\alpha\in\mathbb{Z}[i]$ not in $(1 + i)$ has a unique associate $i^j\alpha$ which is congruent to 1 modulo $(1 + i)^3 = (2 + 2i)$.
(b) Show that the ideal (3) in $\mathbb{Z}[\omega]$, $\omega = (-1 + \sqrt{-3})/2$, is the square of the prime ideal $(\sqrt{-3})$; and that any element $\alpha\in\mathbb{Z}[\omega]$ not in $(\sqrt{-3})$ has a unique associate $(-\omega)^j\alpha$ which is congruent to 1 modulo 3.

19. Consider the elliptic curve $y^2 = x^3 - a$, $a\in\mathbb{F}_q^*$. Recall from Problem 4 of §I.9 that it has $q + 1$ points if $q \equiv 2 \pmod 3$. So suppose that $q \equiv 1 \pmod 3$. Let χ_2 be the nontrivial quadratic character of $\mathbb{F}_q^*$, and let χ_3 be either of the nontrivial characters of $\mathbb{F}_q^*$ of order 3. Prove that the number of $\mathbb{F}_q$-points on the elliptic curve is equal to

$$q + 1 + \chi_2(-a)(\chi_3(a)J(\chi_2, \chi_3) + \bar{\chi}_3(a)J(\chi_2, \bar{\chi}_3)).$$

20. Let $q \equiv 1 \pmod 3$, and let χ_3 be a nontrivial character of $\mathbb{F}_q^*$ of order 3.
(a) Prove that $qJ(\chi_3, \chi_3) = g(\chi_3)^3$.
(b) Prove that $J(\chi_3, \chi_3) \equiv -1 \pmod 3$ in $\mathbb{Z}[\omega]$, where $\omega = (-1 + i\sqrt{3})/2$.
(c) Show that $J(\chi_3, \chi_3) = p$ if $q = p^2$, $p \equiv 2 \pmod 3$.
(d) Suppose that $q = p \equiv 1 \pmod 3$. Choose $a + b\omega$ so that $p = |a + b\omega|^2 = a^2 - ab + b^2$. Show that exactly one of the two ideals $(a + b\omega)$, $(a + b\bar{\omega})$ (without loss of generality, suppose the first one) has the property that

$$\chi_3(x) \equiv x^{(p-1)/3} \pmod{a + b\omega} \quad \text{for all} \quad x\in\mathbb{F}_p.$$

(e) Let $q = p \equiv 1 \pmod 3$, and choose $a + b\omega$ as in part (d). Show that $-J(\chi_3, \chi_3)$ is the unique element generating the ideal $(a + b\omega)$ which is congruent to 1 modulo 3.

21. Let N_r be the number of $\mathbb{F}_{p^r}$-points on the elliptic curve $y^2 = x^3 - a$, where $a\in\mathbb{F}_p^*$, $p \neq 2, 3$.
(a) If $p \equiv 1 \pmod 3$, let χ_2 and χ_3 be nontrivial characters of $\mathbb{F}_p^*$ of order 2 and 3, respectively, and set $\alpha = -\chi_2(-a)\chi_3(4a)J(\chi_3, \chi_3)$. Prove that $N_r = p^r + 1 - \alpha^r - \bar{\alpha}^r$.
(b) If $p \equiv 2 \pmod 3$, let χ_2 and χ_3 be nontrivial characters of $\mathbb{F}_{p^2}^*$ of order 2 and 3, respectively. First prove that χ_2 and χ_3 are both trivial on elements of $\mathbb{F}_p^*$. Now set $\alpha = i\sqrt{p}$. Prove that $N_r = p^r + 1 - \alpha^r - \bar{\alpha}^r$.
(c) Conclude that in both cases the zeta-function is $(1 - 2cT + pT^2)/(1 - T)(1 - pT)$, where $c = 0$ if $p \equiv 2 \pmod 3$, and $c = -\chi_2(-a)\text{Re}(\chi_3(4a)J(\chi_3, \chi_3))$ if $p \equiv 1 \pmod 3$.

22. Let $C \subset \mathbb{P}_K^2$ be the curve $y^2 + ay = x^3$, $a\in K$ (i.e., $\tilde{F}(x, y, z) = y^2z + ayz^2 - x^3$).
(a) Find conditions on the characteristic of K and on $a\in K$ which are equivalent to C being smooth at all of its $K^{\text{alg cl}}$-points.

(b) Let $K = \mathbb{F}_{2^r}$. Show that for r odd, $\#C(\mathbb{F}_{2^r}) = 2^r + 1$ (this is independent of a).

(c) Let $K = \mathbb{F}_{4^r}$, $a \in K$, $a \neq 0$. Let χ_3 be a nontrivial character of K^* of order 3, and let $\bar{\chi}_3$ be the other one. Derive the formula:

$$\#C(\mathbb{F}_{4^r}) = 4^r + 1 + \bar{\chi}_3(a)J(\chi_3, \chi_3) + \chi_3(a)J(\bar{\chi}_3, \bar{\chi}_3).$$

(d) In the situation of part (c), show that $J(\chi_3, \chi_3) = (-1)^{r-1}2^r$; then find a formula for $Z(C/\mathbb{F}_2; T)$ when $a = 1$.

(e) Now let $K = \mathbb{Q}$, $a = 1$. Find a linear change of variables (with coefficients in $\mathbb{Q}$) which transforms C to the elliptic curve $y^2 = x^3 + 16$.

23. Let N_r be the number of $\mathbb{F}_{p^r}$-points on the elliptic curve E_n: $y^2 = x^3 - n^2x$, where $p \nmid 2n$.

(a) Show that if $p \equiv 3 \pmod 4$, then N_r is independent of n; it equals $p^r + 1$ if r is odd; and it equals $(p^{r/2} - (-1)^{r/2})^2$ if r is even.

(b) Now let $p \equiv 1 \pmod 4$. In Problem 8 above, we saw that N_r is independent of n if r is even, and if r is odd it depends only on whether n is a quadratic residue or nonresidue modulo p. For odd r, let N_r^{res} and N_r^{nr} denote the N_r for n a residue and for n a nonresidue, respectively. Show that N_{2r} is a multiple of the least common multiple of N_r^{res} and N_r^{nr}.

(c) For $p = 5$, make a table of N_r^{res} and N_r^{nr} for $r = 1, 3, 5, 7$ and a table of N_r for $r = 2, 4, 6, 8, 10, 12, 14$. In each case, determine the type of the abelian group $E_n(\mathbb{F}_{p^r})$. (See Problems 9 and 11 in §I.9.)

(d) For $p = 13$, make a table of N_r^{res} and N_r^{nr} for $r = 1, 3, 5$ and a table of N_r for $r = 2, 4, 6, 8, 10$; and in each case, find the type of $E_n(\mathbb{F}_{p^r})$.

§3. Varying the prime p

In this section we look at the elliptic curve E_n: $y^2 = x^3 - n^2x$ and its zeta-function $Z(E_n/\mathbb{F}_p; T)$ as p varies. We shall later want to combine these zeta-functions for the various p into a single function, called the Hasse–Weil L-series of the elliptic curve. It is the Hasse–Weil L-function that is intimately related to the group of $\mathbb{Q}$-points on E_n.

The denominator of $Z(E_n/\mathbb{F}_p; T)$ is always $(1 - T)(1 - pT)$. Only the numerator depends on p. If $p|2n$, in which case E_n is not even an elliptic curve, the numerator is simply 1 (see Problem 10 in §II.1). Otherwise, the numerator is a quadratic polynomial in T of the form $(1 - \alpha T)(1 - \bar{\alpha}T)$.

When we later define the Hasse–Weil L-series of E_n, we shall take this quadratic polynomial and replace T by p^{-s} (s is a new complex variable). The resulting expression $(1 - \alpha p^{-s})(1 - \bar{\alpha}p^{-s})$ is called the "Euler factor at p", by analogy with the term in the Euler product expansion of the Riemann zeta-function:

$$\zeta(s) \underset{\text{def}}{=} \sum_{n=1}^{\infty} n^{-s} = \prod_{\text{primes } p} \frac{1}{1 - p^{-s}} \qquad \text{(where Re } s > 1). \tag{3.1}$$

In this section we shall study how this "Euler factor" depends on p. This

dependence will turn out to be described by a certain character χ_n' of $\mathbb{Z}[i]$ (see Problem 9 of the last section).

For the duration of this section we shall let P denote prime ideals of the Gaussian integer ring $\mathbb{Z}[i]$. There are two types: (1) $P = (p)$ for $p \equiv 3 \pmod 4$; (2) $P = (a + bi)$ for $a^2 + b^2 = p \equiv 1 \pmod 4$. In the latter case we have $P\bar{P} = (p)$, and we say that p "splits" in $\mathbb{Z}[i]$. (There is also the special case $P = (1 + i)$, which "ramifies", i.e., $P^2 = (2)$.) The degree of a prime P dividing (p) is defined to be the degree of the field extension $\mathbb{Z}[i]/P$ of $\mathbb{F}_p$; it is 2 in the first case and 1 if p splits. We can then rephrase the theorem in the last section as follows.

Proposition 1.

$$(1 - T)(1 - pT)Z(E_n/\mathbb{F}_p; T) = \prod_{P|(p)} (1 - (\alpha_P T)^{\deg P}), \tag{3.2}$$

where the product is over the (one or two) prime ideals of $\mathbb{Z}[i]$ *dividing* (p), *and where* $\alpha_P = i\sqrt{p}$ *if* $P = (p)$ *and* $\alpha_P = a + bi$ *if* p *splits, where* $a + bi$ *is the unique generator of* P *which is congruent to* $(\frac{n}{p})$ *modulo* $2 + 2i$. *We take* $\alpha_P = 0$ *if* $P|(2n)$.

We now define a map $\tilde{\chi}_n$ on $\mathbb{Z}[i]$ which will be multiplicative and will satisfy $\tilde{\chi}_n(x) = \alpha_P^{\deg P}$ for any generator x of $P = (x)$. This multiplicative map is of the form $\tilde{\chi}_n(x) = x\chi_n'(x)$, where $\chi_n'(x)$ has value 0, ± 1, or $\pm i$. First of all, we define $\chi_n'(x) = 0$ if x has a common factor with $2n$. Next, for $n = 1$ we define $\chi_1'(x)$ to equal i^j, where i^j is the unique power of i such that $i^j x \equiv 1 \pmod{2 + 2i}$. Here x is assumed prime to 2, and hence an element of $(\mathbb{Z}[i]/(2 + 2i))^*$, which has four elements represented by the powers of i. Finally, for other n and for $x \in \mathbb{Z}[i]$ prime to 2, we define $\chi_n'(x) = \chi_1'(x)(\frac{n}{\mathbb{N}x})$, where $\mathbb{N}x = x \cdot \bar{x}$ is a positive odd integer, and $(\frac{\cdot}{m})$ is the Legendre symbol (which extends from prime modulus $(\frac{\cdot}{p})$ to arbitrary positive odd modulus by requiring that $(\frac{n}{m_1 m_2}) = (\frac{n}{m_1})(\frac{n}{m_2})$). To summarize, we have defined:

$$\tilde{\chi}_n(x) = x\chi_n'(x); \qquad \chi_n'(x) = \begin{cases} \chi_1'(x)\left(\dfrac{n}{\mathbb{N}x}\right) & \text{for } x \text{ prime to } 2n; \\ 0 & \text{otherwise}; \end{cases} \tag{3.3}$$

where for x prime to 2

$$\chi_1'(x) = i^j \quad \text{with} \quad i^j x \equiv 1 \pmod{2 + 2i}. \tag{3.4}$$

Suppose that x generates a prime ideal $P = (x)$ not dividing $2n$. If $P = (p)$ with $p \equiv 3 \pmod 4$, then $(\frac{n}{\mathbb{N}x}) = (\frac{n}{p^2}) = 1$, and $\tilde{\chi}_n(x) = i^j x = -p$. That is, $\tilde{\chi}_n$ takes any of the four possible generators of P to α_P^2. If x is any of the four possible generators of a prime P of norm $p \equiv 1 \pmod 4$, then $\tilde{\chi}_n(x) = i^j x(\frac{n}{p}) \equiv (\frac{n}{p}) \pmod{2 + 2i}$, i.e., $\tilde{\chi}_n(x)$ is the unique generator α_P which is congruent to $(\frac{n}{p})$ modulo $2 + 2i$. We have thus shown:

Proposition 2. *The map* $\tilde{\chi}_n$ *defined in* (3.3)–(3.4) *is the unique multiplicative map on* $\mathbb{Z}[i]$ *which coincides with* $\alpha_P^{\deg P}$ *on any generator of a prime ideal* P.

Notice that χ_1' is a character on $(\mathbb{Z}[i]/(2+2i))^*$. It takes any x to the root of unity in the class $1/x$. The general χ_n' is obtained from χ_1' using the Legendre symbol, with the variable x appearing on the bottom. We now use quadratic reciprocity to bring the variable x up on top, thereby showing that χ_n' is a character. At this point recall Problem 9 of the last section, in particular, the definition of a "primitive" character on a number ring.

Proposition 3. *The map* χ_n' *defined in* (3.3)–(3.4) *is a primitive multiplicative character modulo* $(2+2i)n$ *for odd* n *and modulo* $2n$ *for even* n.

PROOF. Suppose x is prime to $2n$. Let $n = 2^\varepsilon l_1 \cdots l_t$, where the l_j are distinct odd prime numbers, and $\varepsilon = 0$ or 1. Note that $\mathbb{N}x$ is a product of odd prime powers, where the primes $p_1, \ldots, p_r$ occurring to odd powers are all congruent to 1 (mod 4). First, it is easy to see that

$$\left(\frac{2}{\mathbb{N}x}\right) = \begin{cases} 1 & \text{if } \mathbb{N}x \equiv 1 \pmod 8; \\ -1 & \text{if } \mathbb{N}x \equiv 5 \pmod 8. \end{cases} \tag{3.5}$$

Next, we compute that

$$\left(\frac{n}{\mathbb{N}x}\right) = \left(\frac{2}{\mathbb{N}x}\right)^\varepsilon \prod_{1\le k\le r,\, 1\le j\le t} \left(\frac{l_j}{p_k}\right) = \left(\frac{2}{\mathbb{N}x}\right)^\varepsilon \prod_{j,k} \left(\frac{p_k}{l_j}\right)$$

by quadratic reciprocity, since $p_k \equiv 1 \pmod 4$. Since $\mathbb{N}x$ is equal to an odd square factor times the product of the p_k, we conclude that

$$\chi_n'(x) = \chi_1'(x)\left(\frac{2}{\mathbb{N}x}\right)^\varepsilon \prod_j \left(\frac{\mathbb{N}x}{l_j}\right) = \chi_1'(x)\left(\frac{2}{\mathbb{N}x}\right)^\varepsilon \left(\frac{\mathbb{N}x}{n_0}\right), \tag{3.6}$$

where $n_0 = n$ if n is odd, $n_0 = n/2$ if n is even.

We now prove the proposition in the case n odd. The proof for n even is very similar, and will be left as an exercise below.

We must first show that $\chi_n'(x)$ depends only on what x is modulo $(2+2i)n$. Suppose that $x' = x + (2+2i)n\beta$. Since $x' \equiv x \pmod{2+2i}$, we clearly have $\chi_1'(x') = \chi_1'(x)$. Next, we have

$$\mathbb{N}x = (x + (2+2i)n\beta)(\bar{x} + (2-2i)n\bar{\beta}) \equiv x \cdot \bar{x} = \mathbb{N}x \pmod n,$$

and hence the Legendre symbols are also equal. (This would not have been clear until we used quadratic reciprocity to bring $\mathbb{N}x$ to the top, obtaining $(\frac{\mathbb{N}x}{n})$ in (3.6).)

To show primitivity, we must show that there is no proper divisor of $(2+2i)n$ such that $\chi_n'(x)$ depends only on what x is modulo that proper divisor. Thus, if χ_n' were not primitive mod $(2+2i)n$, there would exist a prime ideal Q dividing $(2+2i)n$ such that $\chi_n'(x)$ depends only on x modulo

the ideal $((2+2i)n)/Q$. In particular, $\chi_n'(x) \neq -1$ for all $x \equiv 1 \bmod ((2+2i)n)/Q$. We consider three cases, and show that each leads to a contradiction.

(i) $Q = (1+i)$, i.e., $\chi_n'(x) \neq -1$ for all $x = 1 + 2n\beta$, $\beta \in \mathbb{Z}[i]$. But since $\mathbb{N}x \equiv 1 \pmod{n}$, we have $\chi_n'(x) = \chi_1'(x) = \chi_1'(1+2\beta)$, and this value is -1 if, for example, $\beta = i$.

(ii) $Q = (a+bi)$ with $(a+bi)(a-bi) = l \equiv 1 \pmod{4}$, $l \mid n$. Then we are supposing that $\chi_n'(x) \neq -1$ for all x of the form $1 + \beta(2+2i)n(a-bi)/l$, where $\beta \in \mathbb{Z}[i]$. Let $\beta = k(1-i)$, where k is an arbitrary integer, i.e., $x = 1 + 4kn(a-bi)/l$. Then $\chi_1'(x) = 1$, and $\mathbb{N}x \equiv 1 + 8akn/l \pmod{n}$. Hence, $\chi_n'(x) = (\frac{1+8akn/l}{l})$. Since $8an/l$ is prime to l, it follows that $1 + 8akn/l$ runs through all residues modulo l as k varies. In particular, there is a value of k for which $1 + 8akn/l$ is a quadratic nonresidue, i.e., $\chi_n'(x) = -1$, a contradiction.

(iii) $Q = (l)$ with $l \equiv 3 \pmod{4}$. Then we are supposing that $\chi_n'(x) \neq -1$ for $x \equiv 1 \pmod{(2+2i)n/l}$. Since $x \equiv 1 \pmod{2+2i}$, we have $\chi_1'(x) = 1$, and so $\chi_n'(x) = (\frac{\mathbb{N}x}{n}) = (\frac{\mathbb{N}x}{l})$, since $\mathbb{N}x \equiv 1 \pmod{n/l}$. Now since $(2+2i)n/l$ is prime to l, it follows by the Chinese remainder theorem that x of the form $1 + \beta(2+2i)n/l$ runs through all residues of $\mathbb{Z}[i]$ modulo Q. If we consider x modulo Q, i.e., as an element in the field $\mathbb{Z}[i]/Q \approx \mathbb{F}_{l^2}$, then the norm map $x \mapsto \mathbb{N}x = x \cdot \bar{x}$ is simply the norm map from $\mathbb{F}_{l^2}$ to $\mathbb{F}_l$. And the latter map is surjective (for instance, a generator g_2 of $\mathbb{F}_{l^2}^*$ goes to a generator $g = g_2^{l+1}$ of $\mathbb{F}_l^*$). Hence, there are x of the required form for which $\chi_n'(x) = (\frac{\mathbb{N}x}{l}) = -1$. This concludes the proof of the proposition. □

For the remainder of this chapter, we shall let n' denote the conductor of χ_n', i.e., a generator of the largest ideal such that $\chi_n'(x)$ depends only on x modulo that ideal. By Proposition 3, we may choose

$$n' = \begin{cases} (2+2i)n, & n \text{ odd;} \\ 2n, & n \text{ even.} \end{cases} \tag{3.7}$$

Whenever one studies transformation formulas for functions involving characters, as we shall do in the sections that follow, the Gauss sum of the character is almost certain to make an appearance. In preparation for our later derivation of the functional equation for the Hasse—Weil L-series of E_n, we now find a formula for the Gauss sum of the character χ_n': $(\mathbb{Z}[i]/n')^* \mapsto \mathbb{C}^*$ (whose image consists of powers of i).

We define our additive character on $\mathbb{Z}[i]/n'$ by the rule:

$$\psi(x) = e^{2\pi i \operatorname{Re}(x/n')}. \tag{3.8}$$

It is easy to check that ψ is a nontrivial additive character of $\mathbb{Z}[i]/n'$ which satisfies the condition in Problem 9 of the last section, namely, it is nontrivial on the multiples of any proper divisor of n'.

Proposition 4.

$$g(\chi'_n) \underset{\text{def}}{=} \sum_{x \in \mathbb{Z}[i]/n'} \chi'_n(x) e^{2\pi i \operatorname{Re}(x/n')} = \begin{cases} \left(\dfrac{-2}{n}\right) n', & n \text{ odd;} \\ \left(\dfrac{-1}{n_0}\right) in', & n = 2n_0 \text{ even.} \end{cases} \tag{3.9}$$

Proof. To show that $g(\chi'_1) = 2 + 2i$ and $g(\chi'_2) = 4i$ is a short computation that will be left as an exercise (Problem 2 below).

Let m be a positive squarefree odd number. Let $(\frac{}{m})$ denote the character $x \mapsto (\frac{\mathbb{N}x}{m})$ on $(\mathbb{Z}[i]/m)^*$. Then by (3.6) we have

$$\chi'_n = \chi'_1 \cdot \left(\frac{}{n}\right) \quad \text{for } n \text{ odd;} \qquad \chi'_n = \chi'_2 \cdot \left(\frac{}{n_0}\right) \quad \text{for } n = 2n_0 \text{ even.} \tag{3.10}$$

We define the Gauss sum for the character $(\frac{}{m})$ as follows:

$$g\left(\left(\frac{}{m}\right)\right) \underset{\text{def}}{=} \sum_{x \in \mathbb{Z}[i]/m} \left(\frac{\mathbb{N}x}{m}\right) e^{2\pi i \operatorname{Re}(x/m)}. \tag{3.11}$$

We can obtain an alternate form for $g((\frac{}{m}))$ if we replace x by $2x$. (Note that $2x$ runs through $(\mathbb{Z}[i]/m)^*$ as x runs through $(\mathbb{Z}[i]/m)^*$.) Since $\mathbb{N}(2x) = 4\mathbb{N}x$, we have $(\frac{\mathbb{N}2x}{m}) = (\frac{\mathbb{N}x}{m})$. Writing $\operatorname{Re}(2x/m)$ as $\frac{1}{m} \operatorname{Tr} x$, where $\operatorname{Tr} x$ denotes $x + \bar{x}$, we have

$$g\left(\left(\frac{}{m}\right)\right) = \sum_{x \in \mathbb{Z}[i]/m} \left(\frac{\mathbb{N}x}{m}\right) e^{(2\pi i/m) \operatorname{Tr} x}. \tag{3.12}$$

Proposition 4 will follow as an immediate consequence of the following lemmas, which will be proved below.

Lemma 1.

$$g(\chi'_n) = \begin{cases} \left(\dfrac{-2}{n}\right) g(\chi'_1) g\left(\left(\dfrac{}{n}\right)\right), & n \text{ odd;} \\ \left(\dfrac{-1}{n_0}\right) g(\chi'_2) g\left(\left(\dfrac{}{n_0}\right)\right), & n = 2n_0 \text{ even.} \end{cases}$$

Lemma 2. *If* $m = m_1 m_2$, *then* $g\left(\left(\frac{}{m}\right)\right) = g\left(\left(\frac{}{m_1}\right)\right) g\left(\left(\frac{}{m_2}\right)\right)$.

Lemma 3. *If* p *is an odd prime, then* $g\left(\left(\frac{}{p}\right)\right) = p$.

Proof of Lemma 1. First suppose that n is odd. Write x in the form $x = (2 + 2i)x_1 + nx_2$, where x_1 runs through a set of representatives of $\mathbb{Z}[i]$ modulo n and x_2 runs through a set of representatives of $\mathbb{Z}[i]$ modulo $2 + 2i$. By the Chinese remainder theorem, x then runs through a set of

representatives of $\mathbb{Z}[i]$ modulo $(2+2i)n$. By (3.10), we have $\chi_n'(x) = \chi_1'(nx_2)(\frac{\mathbb{N}((2+2i)x_1)}{n})$. By (3.4), we have $\chi_1'(n) = (\frac{-1}{n})$. Also, $\mathbb{N}((2+2i)x_1) = 8\mathbb{N}x_1$, and so the second term becomes $(\frac{2\mathbb{N}x_1}{n})$. Meanwhile, in the additive character we have $\mathrm{Re}(x/n') = \mathrm{Re}(x_1/n + x_2/(2+2i))$. Hence, in the definition (3.9) of $g(\chi_n')$ we have

$$g(\chi_n') = \left(\frac{-1}{n}\right)\left(\frac{2}{n}\right) \sum_{\substack{x_1 \in \mathbb{Z}[i]/n \\ x_2 \in \mathbb{Z}[i]/(2+2i)}} \chi_1'(x_2)\left(\frac{\mathbb{N}x_1}{n}\right) e^{2\pi i\,\mathrm{Re}(x_1/n) + 2\pi i\,\mathrm{Re}(x_2/(2+2i))},$$

and the double sum on the right separates out into $g(\chi_1')g((\frac{}{n}))$.

The proof for even n is very similar, where we write $x = 4x_1 + n_0x_2$. The details will be left as an exercise. □

PROOF OF LEMMA 2. The proof is quite similar to that of Lemma 1. In the definition (3.11) we write $x = x_1m_2 + x_2m_1$, where x_j runs through a set of representatives of $\mathbb{Z}[i]/m_j$, $j = 1, 2$. Since $\mathbb{N}x \equiv m_2^2\mathbb{N}x_1 \pmod{m_1}$ and $\mathbb{N}x \equiv m_1^2\mathbb{N}x_2 \pmod{m_2}$, we have

$$\left(\frac{\mathbb{N}x}{m}\right) = \left(\frac{\mathbb{N}x}{m_1}\right)\left(\frac{\mathbb{N}x}{m_2}\right) = \left(\frac{\mathbb{N}x_1}{m_1}\right)\left(\frac{\mathbb{N}x_2}{m_2}\right).$$

Since also $\mathrm{Re}(x/m) = \mathrm{Re}(x_1/m_1) + \mathrm{Re}(x_2/m_2)$ the sum in (3.11) separates out into a product over x_1 which is equal to $g((\frac{}{m_1}))$ and a product over x_2 which is equal to $g((\frac{}{m_2}))$. □

PROOF OF LEMMA 3. We first consider the case $p \equiv 1 \pmod 4$. Let $p = \beta\cdot\bar{\beta}$, where $\beta = a + bi$. In (3.11), we write $x = x_1\beta + x_2\bar{\beta}$, where x_1 and x_2 each run through $0, 1, 2, \ldots, p-1$ (note that these numbers are representatives of $\mathbb{Z}[i]/\beta$ and also of $\mathbb{Z}[i]/\bar{\beta}$). Again, since β and $\bar{\beta}$ are relatively prime, the Chinese remainder theorem tells us that x will run over $\mathbb{Z}[i]/p$. We have $\mathbb{N}x = (x_1\beta + x_2\bar{\beta})(x_1\bar{\beta} + x_2\beta) \equiv 2x_1x_2\,\mathrm{Re}\,\beta^2 \pmod p$. But $\mathrm{Re}\,\beta^2 = a^2 - b^2 \equiv 2a^2 \pmod p$, since $p = a^2 + b^2$. Thus, since $\mathrm{Re}\,x = x_1a + x_2a$, we have by (3.11)

$$g\left(\left(\frac{}{p}\right)\right) = \sum_{x_1, x_2 \in \mathbb{Z}/p\mathbb{Z}} \left(\frac{ax_1ax_2}{p}\right) e^{(2\pi i/p)(ax_1 + ax_2)}.$$

The double sum separates out into the square of a single sum over $x_1 \in \mathbb{Z}/p\mathbb{Z}$. If we then replace ax_1 by x, we obtain

$$g\left(\left(\frac{}{p}\right)\right) = \left(\sum_{x \in \mathbb{F}_p} \left(\frac{x}{p}\right) e^{2\pi i x/p}\right)^2,$$

which we know equals p by property (2) of Gauss sums for finite fields (see §II.2; also see Problem 4 of §II.2).

Finally, suppose that $p \equiv 3 \pmod 4$. Then (p) is a prime ideal of $\mathbb{Z}[i]$, and $\mathbb{Z}[i]/p$ is the field of p^2 elements. In that case, $g((\frac{}{p}))$ in (3.12) is the Gauss sum for the multiplicative and additive characters of $\mathbb{F}_{p^2}$ obtained

from the multiplicative character $(\frac{x}{p})$ and additive character $e^{2\pi i x/p}$ of $\mathbb{F}_p$ using the norm and the trace. In other words, we are in the situation of the Hasse–Davenport relation (2.10), which tells us that $-g((\frac{\cdot}{p}))$ is the square of the Gauss sum $\Sigma_{x\in\mathbb{F}_p}(\frac{x}{p})e^{2\pi i x/p}$. Again using Problem 4 of §II.2 (this time with $q = p \equiv 3 \pmod 4$), we conclude that $g((\frac{\cdot}{p})) = p$.

This completes the proof of the lemmas, and hence of Proposition 4. □

In Proposition 4, the term $(\frac{-2}{n})$ for n odd, $(\frac{-1}{n_0})$ for n even, is equal to $+1$ if $n \equiv 1, 2, 3 \pmod 8$ and is equal to -1 if $n \equiv 5, 6, 7 \pmod 8$. This sign will turn out to play a crucial role in the functional equation for the Hasse–Weil L-series for E_n. It is called the "root number". If it equals -1, then conjecturally it follows that n must be a congruent number. But there is no known direct reason why any squarefree n congruent to 5, 6, or 7 modulo 8 should be the area of a rational right triangle.

Problems

1. Using (3.6), prove Proposition 3 for $n = 2n_0$ even.
2. Verify the formula in Proposition 4 for $n = 1, 2$ by a direct computation.
3. Prove Lemma 1 for even n.
4. Give another proof of Lemma 3 directly from the definition of $g((\frac{\cdot}{p}))$.

§4. The prototype: the Riemann zeta-function

For Re $s > 1$, the Riemann zeta-function is defined by the convergent infinite sum of reciprocal s-th powers, or alternatively by the product of "Euler factors" $1/(1 - p^{-s})$ with the product over all primes p (see (3.1)). In this section we give a proof of analytic continuation and the functional equation for the Riemann zeta-function $\zeta(s)$. The proof has all of the essential elements that will later be needed to prove analogous facts about the Hasse–Weil L-function of E_n.

We start by recalling some basic tools for working with real- and complex-valued functions. First, we summarize the properties of the gamma-function (for the proofs and further details, see, for example, [Whittaker and Watson 1958, Chapter XII], or [Artin 1964]).

The gamma-function $\Gamma(s)$ interpolates $n!$ in the sense that $\Gamma(n) = (n-1)!$. It can be defined for $s \in \mathbb{C}$ with Re $s > 0$ by the integral

$$\Gamma(s) \underset{\text{def}}{=} \int_0^\infty e^{-t} t^s \frac{dt}{t}. \tag{4.1}$$

It satisfies the relation

$$\Gamma(s+1) = s\Gamma(s), \tag{4.2}$$

which enables one to continue $\Gamma(s)$ analytically onto all of the complex s-plane, except that it has simple poles at $s = 0, -1, -2, -3, \ldots$. The gamma-function also satisfies the relations

$$\Gamma(s)\Gamma(1-s) = \frac{\pi}{\sin(\pi s)} \tag{4.3}$$

and

$$\Gamma\left(\frac{s}{2}\right)\Gamma\left(\frac{s+1}{2}\right) = \sqrt{\pi}2^{1-s}\Gamma(s). \tag{4.4}$$

Finally, using (4.2) and (4.3), one easily sees that the *reciprocal* of the gamma-function is an *entire* function of s.

The gamma-function (4.1) is a special case of a construction known as the "Mellin transform". Given a function $f(t)$ on the positive real axis, its Mellin transform is the function $g(s)$ defined by the formula

$$g(s) = \int_0^\infty f(t)t^s\frac{dt}{t} \tag{4.5}$$

for values of s for which the integral converges. Thus, $\Gamma(s)$ is the Mellin transform of e^{-t}. Notice that for any constant $c > 0$, the Mellin transform of e^{-ct} is $c^{-s}\Gamma(s)$:

$$\int_0^\infty e^{-ct}t^s\frac{dt}{t} = c^{-s}\Gamma(s), \tag{4.6}$$

as we see after a simple change of variables. We shall often have occasion to use (4.6).

Another tool we shall need is the Fourier transform. Let $\mathcal{S}$ be the vector space of infinitely differentiable functions $f\colon \mathbb{R} \mapsto \mathbb{C}$ which decrease at infinity faster than any negative power function, i.e., $|x|^N f(x) \to 0$ as $x \to \pm\infty$ for all N. An example of such a function is $f(x) = e^{-\pi x^2}$. For any $f \in \mathcal{S}$ we define its Fourier transform $\hat{f}$ by:

$$\hat{f}(y) \underset{\text{def}}{=} \int_{-\infty}^{\infty} e^{-2\pi ixy}f(x)dx. \tag{4.7}$$

It is not hard to show that the integral converges for all y, and $\hat{f} \in \mathcal{S}$.

The following properties of the Fourier transform are also easy to verify:

(1) If $a \in \mathbb{R}$ and $g(x) = f(x+a)$, then $\hat{g}(y) = e^{2\pi iay}\hat{f}(y)$.
(2) If $a \in \mathbb{R}$ and $g(x) = e^{2\pi iay}f(x)$, then $\hat{g}(y) = \hat{f}(y-a)$.
(3) If $b > 0$ and $g(x) = f(bx)$, then $\hat{g}(y) = \frac{1}{b}\hat{f}(y/b)$.

For example, to check (3), we compute

$$\hat{g}(y) = \int_{-\infty}^{\infty} e^{-2\pi ixy}f(bx)dx = \int_{-\infty}^{\infty} e^{-2\pi i(x/b)y}f(x)\frac{dx}{b} = \frac{1}{b}\hat{f}(y/b).$$

Proposition 5. *If* $f(x) = e^{-\pi x^2}$, *then* $\hat{f} = f$.

PROOF. Differentiating under the integral sign, we have

$$\hat{f}'(y) = \frac{d}{dy}\int_{-\infty}^{\infty} e^{-2\pi ixy} f(x)dx = -2\pi i\int_{-\infty}^{\infty} e^{-2\pi ixy} xe^{-\pi x^2}dx.$$

Integrating by parts gives

$$\hat{f}'(y) = -2\pi i\, e^{-2\pi ixy}\frac{1}{-2\pi}e^{-\pi x^2}\Big|_{-\infty}^{\infty} + 2\pi i\int_{-\infty}^{\infty} -2\pi iy\, e^{-2\pi ixy}\frac{e^{-\pi x^2}}{-2\pi}dx$$

$$= -2\pi y\int_{-\infty}^{\infty} e^{-2\pi ixy} f(x)dx = -2\pi y \hat{f}(y).$$

Thus, $\hat{f}$ satisfies the differential equation $\hat{f}'(y)/\hat{f}(y) = -2\pi y$; this clearly has solution $\hat{f}(y) = Ce^{-\pi y^2}$, where C is obtained by setting $y = 0$:

$$C = \hat{f}(0) = \int_{-\infty}^{\infty} e^{-\pi x^2}dx = 1.$$

(Recall the evaluation of the latter integral:

$$C^2 = \int_{-\infty}^{\infty} e^{-\pi x^2}dx\int_{-\infty}^{\infty} e^{-\pi y^2}dy = \int_{\mathbb{R}^2} e^{-\pi(x^2+y^2)}dxdy$$

$$= \int_0^{\infty} e^{-\pi r^2}2\pi r\, dr = \int_0^{\infty} e^{-u}du = 1.)$$

Thus, $\hat{f}(y) = e^{-\pi y^2}$, as claimed. □

Proposition 6 (Poisson Summation Formula). *If* $g \in \mathcal{S}$, *then*

$$\sum_{m=-\infty}^{\infty} g(m) = \sum_{m=-\infty}^{\infty} \hat{g}(m). \tag{4.8}$$

PROOF. Define $h(x) = \Sigma_{k=-\infty}^{\infty} g(x + k)$. The function $h(x)$ is periodic with period 1, and has Fourier series $h(x) = \Sigma_{m=-\infty}^{\infty} c_m e^{2\pi imx}$, where

$$c_m = \int_0^1 h(x)e^{-2\pi imx}dx = \int_0^1 \sum_{k=-\infty}^{\infty} g(x+k)e^{-2\pi imx}dx = \int_{-\infty}^{\infty} g(x)e^{-2\pi imx}dx,$$

where we interchanged summation and integration, and made a change of variables (replacing $x + k$ by x) to obtain the last equality. But the last expression is simply $\hat{g}(m)$. Now the left side of (4.8) is $h(0)$, by definition; and the right side is also $h(0)$, as we see by substituting $x = 0$ in the Fourier series for $h(x)$ and using the fact that $c_m = \hat{g}(m)$. □

We now define the theta-function:

$$\theta(t) \underset{\text{def}}{=} \sum_{n=-\infty}^{\infty} e^{-\pi tn^2} \quad \text{for} \quad t > 0. \tag{4.9}$$

Proposition 7. *The theta-function satisfies the functional equation*

$$\theta(t) = \frac{1}{\sqrt{t}}\theta(1/t). \tag{4.10}$$

Proof. We apply Poisson summation to $g(x) = e^{-\pi t x^2}$ for fixed $t > 0$. We write $g(x) = f(\sqrt{t}x)$ with $f(x) = e^{-\pi x^2}$. By Proposition 5 and property (3) of the Fourier transform (with $b = \sqrt{t}$) we have $\hat{g}(y) = t^{-1/2}e^{-\pi y^2/t}$. Then the left side of (4.8) is $\theta(t)$, and the right side is $t^{-1/2}\theta(1/t)$. This proves the proposition. □

We sometimes want to consider $\theta(t)$ for complex t, where we assume that $\operatorname{Re} t > 0$ in the definition (4.9). The functional equation (4.10) still holds for complex t, by the principle of analytic continuation of identities. That is, both sides of (4.10) are analytic functions of t on the right half-plane. Since they agree on the positive real axis, they must be equal everywhere for $\operatorname{Re} t > 0$.

Proposition 8. *As t approaches zero from above, we have*

$$\left|\theta(t) - t^{-1/2}\right| < e^{-C/t} \tag{4.11}$$

for some positive constant C.

Proof. By (4.10) and (4.9), the left side is equal to $2t^{-1/2}\sum_{n=1}^{\infty} e^{-\pi n^2/t}$. Suppose t is small enough so that $\sqrt{t} > 4e^{-1/t}$ and also $e^{-3\pi/t} < \frac{1}{2}$. Then

$$\begin{aligned}\left|\theta(t) - t^{-1/2}\right| &< \tfrac{1}{2}e^{1/t}(e^{-\pi/t} + e^{-4\pi/t} + \cdots) < \tfrac{1}{2}e^{-(\pi-1)/t}(1 + \tfrac{1}{2} + \tfrac{1}{4} + \tfrac{1}{8} + \cdots) \\ &= e^{-(\pi-1)/t}.\end{aligned}$$

Thus, we can take $C = \pi - 1$. □

We now relate $\theta(t)$ to the Riemann zeta-function. Roughly speaking, $\zeta(s)$ is the Mellin transform of $\theta(t)$. The functional equation for $\theta(t)$ then leads us to the functional equation for $\zeta(s)$, and at the same time gives analytic continuation of $\zeta(s)$. We now show how this works.

Theorem. *The Riemann zeta-function $\zeta(s)$ defined by (3.1) for $\operatorname{Re} s > 1$ extends analytically onto the whole complex s-plane, except for a simple pole at $s = 1$ with residue 1. Let*

$$\Lambda(s) \underset{\text{def}}{=} \pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s). \tag{4.12}$$

Then $\Lambda(s)$ is invariant under replacing s by $1 - s$:

$$\Lambda(s) = \Lambda(1 - s).$$

That is, $\zeta(s)$ satisfies the functional equation

$$\pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s) = \pi^{-(1-s)/2}\Gamma\left(\frac{1-s}{2}\right)\zeta(1-s). \tag{4.13}$$

PROOF. Basically, what we want to do is consider the Mellin transform $\int_0^\infty \theta(t)t^s(\frac{dt}{t})$. However, for large t the theta-function is asymptotic to 1 (since all except the $n = 0$ term in (4.9) decrease rapidly); and for t near 0 it looks like $t^{-1/2}$, by Proposition 8. Hence, we must introduce correction terms if we want convergence at both ends. In addition, we replace s by $\frac{s}{2}$ (otherwise, we would end up with $\zeta(2s)$). So we define

$$\phi(s) \underset{\text{def}}{=} \int_1^\infty t^{s/2}(\theta(t) - 1)\frac{dt}{t} + \int_0^1 t^{s/2}\left(\theta(t) - \frac{1}{\sqrt{t}}\right)\frac{dt}{t}. \tag{4.14}$$

In the first integral, the expression $\theta(t) - 1 = 2\sum_{n=1}^\infty e^{-\pi n^2 t}$ approaches zero rapidly at infinity. So the integral converges, and can be evaluated term by term, for *any* s. Similarly, Proposition 8 implies that the second integral converges for any s. In any case, since $\theta(t)$ is bounded by a constant times $t^{-1/2}$ in the interval $(0, 1]$, if we take s with $\operatorname{Re} s > 1$ we can evaluate the second integral as

$$\int_0^1 t^{s/2}\theta(t)\frac{dt}{t} - \int_0^1 t^{(s-1)/2}\frac{dt}{t} = \int_0^1 t^{s/2}\theta(t)\frac{dt}{t} - \frac{2}{s-1}.$$

Thus, for s in the half-plane $\operatorname{Re} s > 1$, we obtain:

$$\begin{aligned}\phi(s) &= 2\sum_{n=1}^\infty \int_1^\infty e^{-\pi n^2 t}t^{s/2}\frac{dt}{t} + \left(\int_0^1 t^{s/2}\frac{dt}{t} + 2\sum_{n=1}^\infty \int_0^1 e^{-\pi n^2 t}t^{s/2}\frac{dt}{t} - \frac{2}{s-1}\right)\\ &= 2\sum_{n=1}^\infty \int_0^\infty e^{-\pi n^2 t}t^{s/2}\frac{dt}{t} + \frac{2}{s} + \frac{2}{1-s}.\end{aligned}$$

Using (4.6) with c replaced by πn^2 and s replaced by $\frac{s}{2}$, we have:

$$\begin{aligned}\frac{1}{2}\phi(s) &= \sum_{n=1}^\infty (\pi n^2)^{-s/2}\Gamma\left(\frac{s}{2}\right) + \frac{1}{s} + \frac{1}{1-s}\\ &= \pi^{-s/2}\Gamma\left(\frac{s}{2}\right)\zeta(s) + \frac{1}{s} + \frac{1}{1-s},\end{aligned} \tag{4.15}$$

where always here $\operatorname{Re} s > 1$.

Now $\phi(s)$ is an entire function of s, since the integrals in (4.13) converge so well for any s, as we saw. Thus, (4.14) shows us that there is a meromorphic function of s on the whole complex plane, namely

$$\frac{\pi^{s/2}}{\Gamma(s/2)}\left(\frac{1}{2}\phi(s) - \frac{1}{s} - \frac{1}{1-s}\right),$$

which is equal to $\zeta(s)$ for $\operatorname{Re} s > 1$. Moreover, since $\pi^{s/2}$, $1/\Gamma(\frac{s}{2})$, and $\phi(s)$ are all entire functions, it follows that the only possible poles are at $s = 0$ and at $s = 1$. But near $s = 0$ we can replace $s\Gamma(\frac{s}{2})$ in the denominator by $2(\frac{s}{2})\Gamma(\frac{s}{2}) = 2\Gamma(\frac{s}{2} + 1)$, which remains nonzero as $s \to 0$. Hence the only pole

is at $s = 1$, where we compute the residue

$$\lim_{s\to 1}(s-1)\frac{\pi^{s/2}}{\Gamma(s/2)}\left(\frac{1}{2}\phi(s) - \frac{1}{s} + \frac{1}{s-1}\right) = \frac{\pi^{1/2}}{\Gamma(1/2)} = 1.$$

It remains to prove the functional equation. Since, by (4.15), $\Lambda(s) = \frac{1}{2}\phi(s) - \frac{1}{s} - \frac{1}{(1-s)}$, and since $\frac{1}{s} + \frac{1}{(1-s)}$ is invariant under replacing s by $1 - s$, it suffices to prove that $\phi(s) = \phi(1 - s)$. This is where we use the functional equation (4.10) for the theta-function. Using (4.10) and replacing t by $\frac{1}{t}$ in (4.14), we obtain (note that $d(\frac{1}{t})/(\frac{1}{t}) = -\frac{dt}{t}$, and $\int_1^\infty$ becomes $\int_1^0 = -\int_0^1$ under the substitution):

$$\phi(s) = \int_0^1 t^{-s/2}\left(\theta\left(\frac{1}{t}\right) - 1\right)\frac{dt}{t} + \int_1^\infty t^{-s/2}\left(\theta\left(\frac{1}{t}\right) - \sqrt{t}\right)\frac{dt}{t}$$

$$\left(\text{replacing } t \text{ by } \frac{1}{t}\right)$$

$$= \int_0^1 t^{-s/2}(\sqrt{t}\theta(t) - 1)\frac{dt}{t} + \int_1^\infty t^{-s/2}(\sqrt{t}\theta(t) - \sqrt{t})\frac{dt}{t} \qquad \text{(by (4.10))}$$

$$= \int_0^1 t^{(1-s)/2}\left(\theta(t) - \frac{1}{\sqrt{t}}\right)\frac{dt}{t} + \int_1^\infty t^{(1-s)/2}(\theta(t) - 1)\frac{dt}{t}$$

$$= \phi(1 - s).$$

This completes the proof of the theorem. □

In a similar way one can prove analytic continuation and a functional equation for the more general series obtained by inserting a Dirichlet character $\chi(n)$ before n^{-s} in (3.1), or, equivalently, inserting $\chi(p)$ before p^{-s} in the Euler product (see Problem 1 below). That is, for any character $\chi\colon (\mathbb{Z}/N\mathbb{Z})^* \to \mathbb{C}^*$, one defines:

$$L(\chi, s) = \sum_{n=1}^{\infty}\frac{\chi(n)}{n^s} = \prod_p \frac{1}{1 - \chi(p)p^{-s}} \qquad \text{(where Re } s > 1). \qquad (4.16)$$

The details of the proof of analytic continuation and the functional equation will be outlined in the form of problems below.

The Hasse–Weil L-function for our elliptic curve E_n, to be defined in the next section, will also turn out to be a series similar to (4.16), except that the summation will be over Gaussian integers x, the denominator will be the norm of x to the s-th power, and the numerator will be $\tilde{\chi}_n(x)$, where $\tilde{\chi}_n$ was defined in (3.3) in the last section. The techniques used in this section to treat the Riemann zeta-function can be modified to give analogous facts—analytic continuation and a functional equation—for the Hasse–Weil L-function for E_n. In the final section we shall use this information to investigate the "critical value" of the Hasse–Weil L-function, which is related to the congruent number problem.

Problems

1. (a) Show that the summation form and the Euler product form of the definition of $L(\chi, s)$ are equal.
   (b) Prove that if χ is nontrivial, then the sum in (4.15) actually converges (conditionally) for $\operatorname{Re} s > 0$.

2. Let $G = \mathbb{Z}/N\mathbb{Z}$, and let $\xi = e^{2\pi i/N}$.
   (a) Define the "finite group Fourier transform" of a function $f\colon G \to \mathbb{C}$ by setting $\hat{f}(a) = \Sigma_{b\in G} f(b)\xi^{-ab}$ for $a \in G$. Prove that $f(b) = \frac{1}{N}\Sigma_{a\in G}\hat{f}(a)\xi^{ab}$.
   (b) For fixed $s \in \mathbb{C}$ with $\operatorname{Re} s > 1$, let $f_s\colon G \to \mathbb{C}$ be the function
$$f_s(b) = \sum_{n\geq 1, n\equiv b \pmod N} n^{-s}.$$
   Prove that for any primitive Dirichlet character χ modulo N and for any s with $\operatorname{Re} s > 1$:
$$L(\chi, s) = \frac{1}{N} g(\chi) \sum_{a\in G} \bar{\chi}(a) \sum_{n=1}^{\infty} \frac{\xi^{-an}}{n^s},$$
   where $g(\chi)$ is the Gauss sum (see Problem 9 of §II.2).
   (c) Take the limit in part (b) as s approaches 1 from above, supposing χ nontrivial. In that way derive a simple formula for $L(\chi, 1)$.
   (d) Define the "dilogarithm" function by $l(x) = \Sigma_{n=1}^{\infty} \frac{x^n}{n^2}$ for $|x| \leq 1$. Express $L(\chi, 2)$ in terms of the dilogarithm.

3. (a) For fixed $t > 0$ and $a \in \mathbb{R}$, what is the Fourier transform of $e^{-\pi t(x+a)^2}$?
   (b) Suppose that $a \in \mathbb{R}$ is in the open interval $(0, 1)$. Define the following functions:
$$\zeta(a, s) = \sum_{n=0}^{\infty} (n + a)^{-s}, \qquad \operatorname{Re} s > 1;$$
$$l(a, s) = \sum_{n=1}^{\infty} n^{-s} e^{2\pi i n a}, \qquad \operatorname{Re} s > 1;$$
$$\theta_a(t) = \sum_{n=-\infty}^{\infty} e^{-\pi t(n+a)^2}, \qquad t > 0;$$
$$\theta^a(t) = \sum_{n=-\infty}^{\infty} e^{2\pi i n a} e^{-\pi t n^2}, \qquad t > 0.$$
   (The notation $l(a, s)$ should not be confused with the function $l(x)$ in Problem 2.) Prove that
   (i) $\theta_a(t) = t^{-1/2}\theta^a(\frac{1}{t})$;
   (ii) $|\theta_a(t) - t^{-1/2}| < e^{-C_1/t}$ as $t \to 0$ for some positive constant C_1;
   (iii) $|\theta^a(t)| < e^{-C_2/t}$ as $t \to 0$ for some positive constant C_2.
   (c) Prove that $\zeta(a, s) + \zeta(1 - a, s)$ as a function of $s \in \mathbb{C}$ extends to a meromorphic function with no pole except for a simple pole at $s = 1$; that the function $l(a, s) + l(1 - a, s)$ extends to an entire function; and that
$$\pi^{-s/2}\Gamma\left(\frac{s}{2}\right)(\zeta(a, s) + \zeta(1 - a, s)) = \pi^{-(1-s)/2}\Gamma\left(\frac{1-s}{2}\right)(l(a, 1 - s) + l(1 - a, 1 - s)).$$

(d) Let χ be a primitive Dirichlet character mod N. Let $L(\chi, s)$ be defined as in (4.16). Prove that for Re $s > 1$:

(i) $\sum_{0<h<N} \chi(h)\zeta\left(\frac{h}{N}, s\right) = N^s L(\chi, s)$;

(ii) $\sum_{0<h<N} \chi(h) l\left(\frac{h}{N}, s\right) = g(\chi) L(\bar{\chi}, s)$.

(e) Suppose that χ is a nontrivial even character, i.e., $\chi(-1) = 1$. Prove that $L(\chi, s)$ extends to an entire function of $s \in \mathbb{C}$, and find a functional equation relating $L(\chi, s)$ to $L(\bar{\chi}, 1 - s)$.

(f) Let χ be a primitive even quadratic character, i.e., $\chi(n) = \pm 1$. Recall from Problem 9(c) of §II.2 that $g(\chi)^2 = N$, so that $g(\chi) = \pm\sqrt{N}$. Suppose you somehow knew that $L(\chi, \frac{1}{2}) \neq 0$. Show that this implies that $g(\chi) = \sqrt{N}$.

(g) With χ as in part (e), show that $L(\chi, s) = 0$ if s is an even negative integer or zero.

(h) With χ as in part (e), express $L'(\chi, -2k)$ in terms of $L(\bar{\chi}, 2k + 1)$. In particular, express $L'(\chi, 0)$ in terms of $L(\bar{\chi}, 1)$.

4. Let χ be a nontrivial even primitive Dirichlet character mod N, and define

$$\theta(\chi, t) = \sum_{n=1}^{\infty} \chi(n) e^{-\pi t n^2} = \frac{1}{2} \sum_{n \in \mathbb{Z}} \chi(n) e^{-\pi t n^2}, \qquad t > 0.$$

(a) Prove that

(i) $\theta(\chi, t) = \frac{1}{2} \sum_{a=1}^{N} \chi(a) \theta_{a/N}(N^2 t)$;

(ii) $\frac{1}{2} \sum_{a=1}^{N} \chi(a) \theta^{a/N}(t) = g(\chi)\theta(\bar{\chi}, t)$;

(iii) $\theta(\chi, t) = \frac{g(\chi)}{\sqrt{N^2 t}} \theta(\bar{\chi}, 1/N^2 t)$.

(b) Show that the Mellin transform of $\theta(\chi, t)$ converges for any s (with no need for any correction term), and that for Re $s > \frac{1}{2}$ it is equal to $\pi^{-s}\Gamma(s)L(\chi, 2s)$.

(c) Use the functional equation in part (a)(iii) to give another proof of the functional equation for $L(\chi, s)$ in Problem 3(e) above.

5. (a) Let $f \in \mathcal{S}$, and $g(x) = f'(x)$. Show that $\hat{g}(y) = 2\pi i y \hat{f}(y)$.

(b) Find the Fourier transform of $(x + a)e^{-\pi t(x+a)^2}$, with t and a as in Problem 3(a).

(c) Let $a \in (0, 1)$. Define $\zeta(a, s)$ and $l(a, s)$ as in Problem 3 above, but now define θ_a and θ^a differently:

$$\theta_a(t) = \sum_{n=-\infty}^{\infty} (n + a) e^{-\pi t(n+a)^2}, \qquad t > 0;$$

$$\theta^a(t) = \sum_{n=-\infty}^{\infty} n e^{2\pi i n a} e^{-\pi t n^2}, \qquad t > 0.$$

Prove that:

(i) $\theta_a(t) = -it^{-3/2}\theta^a(\frac{1}{t})$;

(ii) $|\theta_a(t)| < e^{-C_1/t}$ as $t \to 0$ for some positive constant C_1;

(iii) $|\theta^a(t)| < e^{-C_2/t}$ as $t \to 0$ for some positive constant C_2.

(d) Express the Mellin transform of θ_a and θ^a in terms of $\zeta(a, s)$ and $l(a, s)$; prove that $\zeta(a, s) - \zeta(1 - a, s)$ and $l(a, s) - l(1 - a, s)$ extend to entire functions of $s \in \mathbb{C}$; and derive a functional equation relating these two functions.

(e) Suppose that χ is a primitive odd character mod N, i.e., $\chi(-1) = -1$. Prove that $L(\chi, s)$ extends to an entire function of $s \in \mathbb{C}$, and find a functional equation relating $L(\chi, s)$ to $L(\bar{\chi}, 1 - s)$.

(f) Let χ be an odd quadratic character. Show that if you somehow knew that $L(\chi, 1/2) \neq 0$, then this would imply that $g(\chi) = i\sqrt{N}$ (rather than $-i\sqrt{N}$).

(g) With χ as in part (e), show that $L(\chi, s) = 0$ if s is a negative odd integer.

(h) With χ as in part (e), express $L'(\chi, 1 - 2k)$ in terms of $L(\bar{\chi}, 2k)$. In particular, express $L'(\chi, -1)$ in terms of the dilogarithm. For example, express $L'(\chi, -1)$, where $\chi(n) = (\frac{n}{3})$, in terms of the dilogarithm.

6. Let χ be an odd primitive Dirichlet character mod N, and define

$$\theta(\chi, t) = \sum_{n=1}^{\infty} n\chi(n)e^{-\pi t n^2} = \frac{1}{2}\sum_{n \in \mathbb{Z}} n\chi(n)e^{-\pi t n^2}, \qquad t > 0.$$

(Note that this is different from the definition of $\theta(\chi, t)$ for even χ in Problem 4.) Let θ_a and θ^a be as in Problem 5(c).

(a) Prove that:

(i) $\theta(\chi, t) = \dfrac{N}{2}\displaystyle\sum_{a=1}^{N} \chi(a)\theta_{a/N}(N^2 t)$;

(ii) $\dfrac{1}{2}\displaystyle\sum_{a=1}^{N} \chi(a)\theta^{a/N}(t) = g(\chi)\theta(\bar{\chi}, t)$;

(iii) $\theta(\chi, t) = -iN^{-2}t^{-3/2}g(\chi)\theta(\bar{\chi}, 1/N^2 t)$.

(b) Show that the Mellin transform of $\theta(\chi, t)$ converges for any s, and that for $\operatorname{Re} s > \frac{1}{2}$ it is equal to $\pi^{-s}\Gamma(s)L(\chi, 2s - 1)$.

(c) Use the functional equation in part (a)(iii) to give another proof of the functional equation for $L(\chi, s)$ in Problem 5(e) above.

7. Let χ be the character mod 12 such that $\chi(\pm 1) = 1$, $\chi(\pm 5) = -1$. Let $\eta(z) \underset{\text{def}}{=} \theta(\chi, -iz/12)$ for $\operatorname{Im} z > 0$. Prove that $\eta(-1/z) = \sqrt{z/i}\,\eta(z)$, where we take the branch of $\sqrt{z/i}$ which has value 1 when $z = i$. We shall later encounter $\eta(z)$ again, and give a different expression for it and a different proof of its functional equation.

8. (a) Use the functional equations derived above to express $l(a, 1 - s)$ in terms of $\zeta(a, s)$ and $\zeta(1 - a, s)$.

(b) Use the properties (4.3) and (4.4) of the gamma-function along with part (a) to show that

$$l(a, 1 - s) = \Gamma(s)(2\pi)^{-s}e^{is\pi/2}(\zeta(a, s) + e^{-is\pi}\zeta(1 - a, s)).$$

(c) For $a \in \mathbb{C}$, $a \neq -n$, define $(a + n)^{-s}$ to mean $e^{-s\log(a+n)}$, where we take the branch of log having imaginary part in $(-\pi, \pi]$. Show that for $a \in \mathbb{C}$, $\operatorname{Im} a > 0$, $\operatorname{Re} s > 1$, one has:

$$l(a, 1 - s) = \Gamma(s)(2\pi)^{-s}e^{is\pi/2}\sum_{n=-\infty}^{\infty}(a + n)^{-s}.$$

(d) Let $s = k$ be a positive even integer. Show that for $a \in \mathbb{C}$, $\operatorname{Im} a > 0$:

$$\sum_{n=-\infty}^{\infty} \frac{1}{(a+n)^k} = \frac{(2\pi i)^k}{(k-1)!} \sum_{n=1}^{\infty} n^{k-1} e^{2\pi i n a}.$$

(e) Give a second derivation of the formula in part (d) by successively differentiating the formula

$$\pi \cot(\pi a) = \frac{1}{a} + \sum_{n=1}^{\infty} \frac{1}{a+n} + \frac{1}{a-n}.$$

§5. The Hasse–Weil L-function and its functional equation

Earlier in this chapter we studied the congruence zeta-function $Z(E/\mathbb{F}_p; T)$ for our elliptic curves E_n: $y^2 = x^3 - n^2 x$. That function was defined by a generating series made up from the number $N_r = N_{r,p}$ of $\mathbb{F}_{p^r}$-points on the elliptic curve reduced mod p. We now combine these functions for all p to obtain a function which incorporates the numbers $N_{r,p}$ for all possible prime powers p^r, i.e., the numbers of points on E_n over all finite fields.

Let s be a complex variable. We make the substitution $T = p^{-s}$ in $Z(E_n/\mathbb{F}_p; T)$, and define the Hasse–Weil L-function $L(E_n, s)$ as follows:

$$L(E_n, s) \underset{\text{def}}{=} \frac{\zeta(s)\zeta(s-1)}{\prod_p Z(E_n/\mathbb{F}_p; p^{-s})}$$

$$= \prod_{p \nmid 2n} \frac{1}{1 - 2a_{E_n,p} p^{-s} + p^{1-2s}} \tag{5.1}$$

$$= \prod_{P \nmid 2n} \frac{1}{1 - \alpha_P^{\deg P} (\mathbb{N}P)^{-s}}. \tag{5.2}$$

We must first explain the meaning of these products, why they are equivalent, and what restriction on $s \in \mathbb{C}$ will ensure convergence. In (5.1) we are using the form of the congruence zeta-function in the theorem in §II.2 (see the first equality in (2.7)), where the notation $a_{E_n,p}$ indicates that the coefficient a depends on E_n and also on the prime p. We put the term $\zeta(s)\zeta(s-1)$ in the definition so that the uninteresting part of the congruence zeta-function—its denominator—disappears, as we see immediately by replacing $\zeta(s)$ and $\zeta(s-1)$ by their Euler products (see (3.1)). Note that when $p|2n$, the denominator term is all there is (see Problem 10 of §II.1), so we only have a contribution of 1 to the product in that case; so those primes do not appear in the product in (5.1).

In (5.2) the product is over all prime ideals P of $\mathbb{Z}[i]$ which divide primes p of good reduction. Recall that those primes are of two types: $P = (p)$, $p \equiv 3 \pmod 4$, $\deg P = 2$, $\mathbb{N}P = p^2$; and $P = (a+bi)$, $a^2 + b^2 = p \equiv 1 \pmod 4$, $\deg P = 1$, $\mathbb{N}P = p$. The meaning of α_P and the equivalence of (5.1) and (5.2) are contained in Proposition 1 (see (3.2)).

As in the case of the Riemann zeta-function, we can expand the Euler product, writing each term as a geometric series and multiplying all of the geometric series corresponding to each prime. The result is a Dirichlet series, i.e., a series of the form

$$L(E_n, s) = \sum_{m=1}^{\infty} b_{m,n} m^{-s}. \tag{5.3}$$

Before discussing the "additive" form of $L(E_n, s)$ in detail, let us work out the values of the first few $b_{m,n}$ for the example of the elliptic curve $E_1: y^2 = x^3 - x$. We first compute the first few values of $a_{E_1,p}$ in (5.1). If $p \equiv 3 \pmod 4$, then $a_{E_1,p} = 0$. If $p \equiv 1 \pmod 4$, there are two easy ways to compute $a = a_{E_1,p}$: (1) as the solution to $a^2 + b^2 = p$ for which $a + bi \equiv 1 \pmod{2 + 2i}$; (2) after counting the number N_1 of $\mathbb{F}_p$-points on E_1, we have $2a = p + 1 - N_1$ (see (1.5)). Here is the result:

$$\begin{aligned} L(E_1, s) &= \frac{1}{1 + 3\cdot 9^{-s}} \cdot \frac{1}{1 + 2\cdot 5^{-s} + 5\cdot 25^{-s}} \cdot \frac{1}{1 + 7\cdot 49^{-s}} \cdot \frac{1}{1 + 11\cdot 121^{-s}} \\ &\quad \cdot \frac{1}{1 - 6\cdot 13^{-s} + 13\cdot 169^{-s}} \cdot \frac{1}{1 - 2\cdot 17^{-s} + 17\cdot 289^{-s}} \cdots \\ &= 1 - 2\cdot 5^{-s} - 3\cdot 9^{-s} + 6\cdot 13^{-s} + 2\cdot 17^{-s} + \sum_{m \geq 25} b_{m,n} m^{-s}. \end{aligned} \tag{5.4}$$

We have not yet discussed convergence of the series or product for $L(E_n, s)$. Using (5.2) and the standard criterion for an infinite product to converge to a nonzero value, we are led to consider $\Sigma_P |\alpha_P|^{\deg P} (\mathbb{N}P)^{-s}$ for s real. By Proposition 1, we have $|\alpha_P|^{\deg P} = \mathbb{N}P^{1/2}$. In addition, $\mathbb{N}P^{1/2-s} \leq p^{1/2-s}$ for $s \geq \frac{1}{2}$ (where $P = (p)$ or else $P\bar{P} = (p)$). Since there are at most two P's for each p, it follows that the sum is bounded by $2\,\Sigma_p p^{1/2-s}$, which converges if $\operatorname{Re} s > \frac{3}{2}$. To summarize, the right half-plane of guaranteed convergence is 1/2 to the right of the right half-plane of convergence for the Riemann zeta-function, because we have a term of absolute value $\sqrt{p}$ in the Euler product which was absent in the case of $\zeta(s)$.

We now discuss the additive form of $L(E_n, s)$ in more detail. Using Proposition 2, we can rewrite (5.2) in terms of the map $\tilde{\chi}_n$ defined in (3.3)–(3.4):

$$L(E_n, s) = \prod_{P \nmid 2n} \left(1 - \frac{\tilde{\chi}_n(P)}{(\mathbb{N}P)^s}\right)^{-1}, \tag{5.5}$$

where we have used $\tilde{\chi}_n(P)$ to denote its value at any generator of the ideal P. Notice that, since $\tilde{\chi}_n$ is a multiplicative map taking the value 1 at all four units ± 1, $\pm i$, we may regard it equally well as a map on elements x of $\mathbb{Z}[i]$ or on ideals I.

We can now expand the product (5.5) in the same way one does for the Riemann zeta-function and for Dirichlet L-series (see Problem 1(a) in the last section). We use the two facts: (1) every ideal I has a unique factorization

as a product of prime power ideals; and (2) both $\tilde{\chi}_n$ and $\mathbb{N}$ are multiplicative: $\tilde{\chi}_n(I_1 I_2) = \tilde{\chi}_n(I_1)\tilde{\chi}_n(I_2)$, $\mathbb{N}(I_1 I_2) = \mathbb{N}I_1 \cdot \mathbb{N}I_2$. Then, by multiplying out the geometric series, we obtain:

$$L(E_n, s) = \sum_I \tilde{\chi}_n(I)(\mathbb{N}I)^{-s}, \tag{5.6}$$

where the sum is over all nonzero ideals of $\mathbb{Z}[i]$.

A series of the form (5.6) is called a "Hecke L-series", and the map $\tilde{\chi}_n$ is an example of a "Hecke character". In a Hecke L-series, the sum on the right of (5.6) is taken over all nonzero ideals in some number ring. A multiplicative map χ on the ideals in that ring is said to be a Hecke character if the following condition holds. There is some fixed ideal $\mathfrak{f}$ and a fixed set of integers n_σ, one for each imbedding σ of the number field into $\mathbb{Q}^{\text{alg cl}}$, such that if I is a principal ideal generated by an element x which is congruent to 1 modulo the ideal $\mathfrak{f}$, then $\chi(I) = \Pi_\sigma \sigma(x)^{n_\sigma}$. In our example, the number ring is $\mathbb{Z}[i]$; there are two imbeddings $\sigma_1 =$ identity, $\sigma_2 =$ complex conjugation in $\text{Gal}(\mathbb{Q}[i]/\mathbb{Q})$; we take $n_{\sigma_1} = 1$, $n_{\sigma_2} = 0$; and we take $\mathfrak{f} = (n')$ ($n' = (2 + 2i)n$ if n is odd, $2n$ if n is even). Then the condition simply states that $\tilde{\chi}_n((x)) = x$ if $x \equiv 1 \pmod{n'}$.

It is very useful when the Hasse–Weil L-series of an elliptic curve turns out to be a Hecke L-series. In that case one can work with it much as with Dirichlet L-series, for example, proving analytic continuation and a functional equation. It can be shown that the Hasse–Weil L-series of an elliptic curve with complex multiplication (see Problem 8 of §I.8) is always a Hecke L-series.

The relation between the additive form (5.6) and the additive form (5.3) is quite simple. We obtain (5.3) by collecting all terms corresponding to ideals I with the same norm, i.e.,

$$b_{m,n} = \sum_{I \text{ with } \mathbb{N}I = m} \tilde{\chi}_n(I).$$

Notice that, since $\tilde{\chi}_n(I) = \tilde{\chi}_1(I) \cdot (\frac{n}{\mathbb{N}I})$ by (3.3), we have

$$b_{m,n} = \left(\frac{n}{m}\right) \sum_{I \text{ with } \mathbb{N}I = m} \tilde{\chi}_1(I) = \left(\frac{n}{m}\right) b_m,$$

where we have denoted $b_m = b_{m,1}$. Thus, if for fixed n we let χ_n denote the multiplicative map on $\mathbb{Z}$ given by $m \mapsto (\frac{n}{m})$ (for m prime to $2n$), we have

$$\begin{aligned} L(E_n, s) &= \sum_{m=1}^{\infty} \chi_n(m) b_m m^{-s} \\ &= 1 - 2\left(\frac{n}{5}\right)5^{-s} - 3\left(\frac{n}{3}\right)^2 9^{-s} + 6\left(\frac{n}{13}\right)13^{-s} + 2\left(\frac{n}{17}\right)17^{-s} + \cdots \end{aligned} \tag{5.7}$$

(note: $(\frac{n}{3})^2$ is 1 if $3 \nmid n$ and 0 if $3|n$); one says that $L(E_n, s)$ is a "twisting" of $L(E_1, s) = \Sigma\, b_m m^{-s}$ by the character χ_n. One can verify that for n squarefree, the conductor of χ_n is n when $n \equiv 1 \pmod 4$ and is $4n$ when $n \equiv 2$ or 3

mod 4 (this follows from quadratic reciprocity). In other words, χ_n is a primitive Dirichlet character modulo n or $4n$.

To keep the notation clear in our minds, let us review the meaning of χ_n, χ_n', and $\tilde{\chi}_n$. First, χ_n is a map from $\mathbb{Z}$ to $\{\pm 1, 0\}$ which is defined by the Legendre symbol on integers prime to $2n$. Second, χ_n' is a map from $\mathbb{Z}[i]$ to $\{\pm 1, \pm i, 0\}$ which takes elements x prime to $2n$ to the unique power of i such that $\chi_n'(x)x \equiv \chi_n(\mathbb{N}x)$ modulo $2 + 2i$ (see (3.3)–(3.4)). Thirdly, $\tilde{\chi}_n$ is a map from $\mathbb{Z}[i]$ to $\mathbb{Z}[i]$ which takes an element x to $x\chi_n'(x)$; also, $\tilde{\chi}_n$ can be regarded as a map from ideals of $\mathbb{Z}[i]$ to elements of $\mathbb{Z}[i]$ which takes an ideal I prime to $2n$ to the unique generator of I which is congruent to $\chi_n(\mathbb{N}I)$ modulo $2 + 2i$.

The character χ_n is intimately connected with the quadratic field $\mathbb{Q}(\sqrt{n})$. Namely, if $m = p \neq 2$ is a prime number, then the value of $\chi_n(p) = (\frac{n}{p})$ shows whether p splits into a product of two prime ideals $(p) = P_1P_2$ in $\mathbb{Q}(\sqrt{n})$ (this happens if $(\frac{n}{p}) = 1$), remains prime (if $(\frac{n}{p}) = -1$), or ramifies $(p) = P^2$ (if $(\frac{n}{p}) = 0$, i.e., $p|n$). (See [Borevich and Shafarevich 1966].) We say that χ_n is the quadratic character associated to the field $\mathbb{Q}(\sqrt{n})$.

It is not surprising that the character corresponding to the field $\mathbb{Q}(\sqrt{n})$ appears in the formula (5.7) which links $L(E_n, s)$ with $L(E_1, s)$. In fact, if we allow ourselves to make a linear change of variables *with coefficients in* $\mathbb{Q}(\sqrt{n})$, then we can transform E_n: $y^2 = x^3 - n^2x$ to E_1: $y'^2 = x'^3 - x'$ by setting $y = n\sqrt{n}y'$, $x = nx'$. One says that E_n and E_1 are isomorphic "over the field $\mathbb{Q}(\sqrt{n})$."

Returning now to the expression (5.6) for $L(E_n, s)$, we see that it can also be written as a sum over elements of $\mathbb{Z}[i]$ rather than ideals. We simply note that every nonzero ideal has four generators, and so appears four times if we list elements instead of ideals. Thus,

$$b_{m,n} = \frac{1}{4} \sum_{a+bi \text{ with } a^2+b^2=m} \tilde{\chi}_n(a + bi),$$

and

$$\begin{aligned} L(E_n, s) &= \frac{1}{4} \sum_{x \in \mathbb{Z}[i]} \tilde{\chi}_n(x)(\mathbb{N}x)^{-s} \\ &= \frac{1}{4} \sum_{a+bi \in \mathbb{Z}[i]} \frac{(a + bi)\chi_n'(a + bi)}{(a^2 + b^2)^s}, \end{aligned} \tag{5.8}$$

where χ_n' was defined in (3.3)–(3.4). (The sums are over nonzero x, $a + bi$.)

Notice the analogy between the sum (5.8) and Dirichlet L-series. The only differences are that the number ring is $\mathbb{Z}[i]$ rather than $\mathbb{Z}$, and our Hecke character $\tilde{\chi}_n(x)$ includes an ordinary character $\chi_n'(x)$ (with values in the roots of unity) multiplied by x.

We now proceed to show that $L(E_n, s)$ can be analytically continued to the left of $\operatorname{Re} s = \frac{3}{2}$, in fact, to an entire function on the whole complex plane; and that it satisfies a functional equation relating $L(E_n, s)$ to $L(E_n, 2 - s)$.

Since $L(E_n, s)$ is a "two-dimensional" sum over $\mathbb{Z}[i] \approx \mathbb{Z}^2$, i.e., over pairs of integers rather than integers, it follows that we shall need to look at Fourier transforms, the Poisson summation formula, and theta-functions in two variables. We shall give the necessary ingredients as a sequence of propositions whose proofs are no harder than the analogous results we proved in the last section for the case of one variable.

Since the definitions and properties we need in two dimensions are just as easy to state and prove in n dimensions, we shall consider functions on $\mathbb{R}^n$. For now, n will denote the number of variables (not to be confused with our use of n when writing E_n: $y^2 = x^3 - n^2x$, χ_n, etc.). We will use $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$ to denote vectors in $\mathbb{R}^n$. As usual, we let $x \cdot y = x_1y_1 + \cdots + x_ny_n$, $|x| = \sqrt{x \cdot x}$. We shall also use the dot-product notation when the vectors are in $\mathbb{C}^n$; for example, if $n = 2$ we have $x \cdot (1, i) = x_1 + x_2 i$.

Let $\mathcal{S}$ be the vector space of functions $f\colon \mathbb{R}^n \to \mathbb{C}$ which are bounded, smooth (i.e., all partial derivatives exist and are continuous), and rapidly decreasing (i.e., $|x|^N f(x)$ approaches zero whenever $|x|$ approaches infinity for any N). For $f \in \mathcal{S}$ we define the Fourier transform $\hat{f}\colon \mathbb{R}^n \to \mathbb{C}$ as follows (where dx denotes $dx_1 dx_2 \cdots dx_n$):

$$\hat{f}(y) = \int_{\mathbb{R}^n} e^{-2\pi i x \cdot y} f(x)\,dx. \tag{5.9}$$

This integral converges for all $y \in \mathbb{R}^n$, and $\hat{f} \in \mathcal{S}$.

Proposition 9. *Let $f\colon \mathbb{R}^n \to \mathbb{C}$, $g\colon \mathbb{R}^n \to \mathbb{C}$ be functions in $\mathcal{S}$.*

(1) *If $a \in \mathbb{R}^n$ and $g(x) = f(x + a)$, then $\hat{g}(y) = e^{2\pi i a \cdot y}\hat{f}(y)$.*
(2) *If $a \in \mathbb{R}^n$ and $g(x) = e^{2\pi i a \cdot x} f(x)$, then $\hat{g}(y) = \hat{f}(y - a)$.*
(3) *If $b \in \mathbb{R}$, $b > 0$, and $g(x) = f(bx)$, then $\hat{g}(y) = b^{-n}\hat{f}(y/b)$.*
(4) *If $f(x) = e^{-\pi x \cdot x}$, then $\hat{f} = f$.*

Proposition 10 (Poisson Summation Formula). *If $g \in \mathcal{S}$, then*

$$\sum_{m \in \mathbb{Z}^n} g(m) = \sum_{m \in \mathbb{Z}^n} \hat{g}(m).$$

The proofs of Propositions 9 and 10 are completely similar to those of properties (1)–(3) of the Fourier transform in one variable and Propositions 5 and 6 of the last section. One simply has to proceed one variable at a time.

If $w \in \mathbb{C}^n$ and $f \in \mathcal{S}$, we let $w \cdot \frac{\partial}{\partial x} f \underset{\text{def}}{=} w_1 \frac{\partial f}{\partial x_1} + w_2 \frac{\partial f}{\partial x_2} + \cdots + w_n \frac{\partial f}{\partial x_n}$.

Proposition 11. *If $f \in \mathcal{S}$ and $g = w \cdot \frac{\partial}{\partial x} f$, then $\hat{g}(y) = 2\pi i w \cdot y \hat{f}(y)$.*

Proof. Since both sides of the equality are linear in w, it suffices to prove the proposition when w is the j-th standard basis vector, i.e., to prove that the Fourier transform of $\frac{\partial}{\partial x_j} f(x)$ is $2\pi i y_j \hat{f}(y)$. This is easily done by sub-

stituting $\frac{\partial}{\partial x_j} f(x)$ in place of $f(x)$ in (5.9) and integrating by parts with respect to the j-th variable (see Problem 5(a) in the last section). □

For the rest of this section, we take $n = 2$ in Propositions 9–11, and we return to our earlier use of the letter n in E_n, χ_n, etc.

Theorem. *The Hasse–Weil L-function $L(E_n, s)$ for the elliptic curve E_n: $y^2 = x^3 - n^2 x$, which for* $\operatorname{Re} s > \frac{3}{2}$ *is defined by* (5.1), *extends analytically to an entire function on the whole complex s-plane. In addition, let*

$$N = 4|n'|^2 = \begin{cases} 32n^2, & n \text{ odd;} \\ 16n^2, & n \text{ even.} \end{cases} \tag{5.10}$$

Let

$$\Lambda(s) \underset{\text{def}}{=} \left(\frac{\sqrt{N}}{2\pi}\right)^s \Gamma(s) L(E_n, s). \tag{5.11}$$

Then $L(E_n, s)$ satisfies the functional equation

$$\Lambda(s) = \pm\Lambda(2 - s), \tag{5.12}$$

where the "root number" ± 1 is equal to 1 *if* $n \equiv 1, 2, 3 \pmod 8$ *and is equal to* -1 *if* $n \equiv 5, 6, 7 \pmod 8$.

PROOF. The proof is closely parallel to the proof of analytic continuation and the functional equation for Dirichlet L-series with odd character, which was outlined in Problem 5 of the last section. Namely, we express $L(E_n, s)$, written in the form (5.8), in terms of the Mellin transform of a two-dimensional version of the theta-function $\theta_a(t)$ defined in Problem 5(c). We shall use the letter u rather than a to avoid confusion with the use of a in (5.8).

Thus, let $u = (u_1, u_2) \in \mathbb{R}^2$, where $u \notin \mathbb{Z}^2$, and let $t \in \mathbb{R}$ be positive. Let w be the fixed vector $(1, i) \in \mathbb{C}^2$, so that, for example, $m \cdot w = m_1 + m_2 i$ for $m \in \mathbb{Z}^2$. We define:

$$\theta_u(t) = \sum_{m \in \mathbb{Z}^2} (m + u) \cdot w\, e^{-\pi t |m+u|^2}; \tag{5.13}$$

$$\theta^u(t) = \sum_{m \in \mathbb{Z}^2} m \cdot w\, e^{2\pi i m \cdot u} e^{-\pi t |m|^2}. \tag{5.14}$$

Regarding u and t as fixed, we find a functional equation for $\theta_u(t)$ by means of the Poisson summation formula (Proposition 10); to obtain $\theta_u(t)$ on the left side of Proposition 10, we choose

$$g(x) = (x + u) \cdot w\, e^{-\pi t |x+u|^2}. \tag{5.15}$$

To find the Fourier transform of $g(x)$, and hence the right side of the Poisson summation formula, we proceed in several steps, writing $f(x) = e^{-\pi |x|^2}$, $g_1(x) = f(\sqrt{t}x)$, $g_2(x) = w \cdot \frac{\partial}{\partial x} g_1(x)$, and finally $g(x) = \frac{-1}{2\pi t} g_2(x + u)$. We have:

$$\hat{f}(y) = e^{-\pi|y|^2} \text{ by Proposition 9, part (4);}$$
$$\hat{g}_1(y) = t^{-1}e^{-(\pi/t)|y|^2} \text{ by Proposition 9, part (3);}$$
$$\hat{g}_2(y) = 2\pi i t^{-1} w \cdot y\, e^{-(\pi/t)|y|^2} \text{ by Proposition 11;}$$
$$\hat{g}(y) = -it^{-2} w \cdot y\, e^{2\pi i u, y} e^{-(\pi/t)|y|^2} \text{ by Proposition 9, part (1).}$$

If we now evaluate $\hat{g}(m)$ for $m \in \mathbb{Z}^2$, and sum over all m, we obtain the functional equation

$$\theta_u(t) = \frac{-i}{t^2}\theta^u\left(\frac{1}{t}\right). \tag{5.16}$$

We now consider the Mellin transform of $\theta_u(t)$: $\int_0^\infty t^s\theta_u(t)\frac{dt}{t}$, and show that the integral converges to an entire function of s. First, for large t it is easy to bound the integrand by something of the form e^{-ct}, using the fact that $|m + u|^2$ is bounded away from zero, since u is not in $\mathbb{Z}^2$. Next, for t near zero one uses the functional equation (5.16) and a bound for $\theta^u(\frac{1}{t})$ of the form $e^{-c/t}$, where we use the fact that the only term in (5.14) with $|m|^2 = 0$ vanishes because of the factor $m \cdot w$. These bounds make it a routine matter to show that the integral converges for all s, and that the Mellin transform is analytic in s.

If we now take $\operatorname{Re} s > \frac{3}{2}$, we can evaluate the Mellin transform integral term by term, obtaining a sum that begins to look like our L-function:

$$\int_0^\infty t^s\theta_u(t)\frac{dt}{t} = \sum_{m\in\mathbb{Z}^2}(m+u)\cdot w\int_0^\infty t^s e^{-\pi t|m+u|^2}\frac{dt}{t}$$
$$= \pi^{-s}\Gamma(s)\sum_{m\in\mathbb{Z}^2}\frac{(m+u)\cdot w}{|m+u|^{2s}} \qquad \text{(see (4.6)).}$$

Now for $\operatorname{Re} s > \frac{3}{2}$, we can rewrite $L(E_n, s)$ as a linear combination of these sums with various u.

We now suppose that n is odd. The case $n = 2n_0$ even is completely similar, and will be left as an exercise below. We take $w = (1, i)$. If we use (5.8) and recall that $\chi_n'(x)$ depends only on x modulo $n' = (2 + 2i)n$, and hence, *a fortiori*, only on x modulo $4n$, we obtain:

$$L(E_n, s) = \frac{1}{4}\sum_{0\le a,b<4n}\chi_n'(a+bi)\sum_{m\in\mathbb{Z}^2}\frac{a+bi+4nm\cdot w}{|(a,b)+4nm|^{2s}},$$
$$= \frac{1}{4}(4n)^{1-2s}\sum_{0\le a,b<4n}\chi_n'(a+bi)\sum_{m\in\mathbb{Z}^2}\frac{(m+(\frac{a}{4n},\frac{b}{4n}))\cdot w}{|m+(\frac{a}{4n},\frac{b}{4n})|^{2s}}.$$

Thus

$$\pi^{-s}\Gamma(s)L(E_n, s) = \frac{1}{4}(4n)^{1-2s}\sum_{\substack{0\le a,b<4n\\(a,b)\ne(0,0)}}\chi_n'(a+bi)\int_0^\infty t^s\theta_{a/4n,b/4n}(t)\frac{dt}{t}. \tag{5.17}$$

Since the integral inside the finite sum is an entire function of s, as are the

functions $(4n)^{1-2s}$ and $\pi^s/\Gamma(s)$, we conclude that $L(E_n, s)$ has an analytic continuation to an entire function of s.

Moreover, we can transform this integral using the functional equation (5.16) and replacing t by $\frac{1}{t}$:

$$\int_0^\infty t^s \theta_{a/4n,b/4n}(t)\frac{dt}{t} = -i\int_0^\infty t^{s-2}\theta^{a/4n,b/4n}\left(\frac{1}{t}\right)\frac{dt}{t} = -i\int_0^\infty t^{2-s}\theta^{a/4n,b/4n}(t)\frac{dt}{t}.$$

In the entire function (5.17) we now suppose that $\operatorname{Re} 2 - s > \frac{3}{2}$ (i.e., $\operatorname{Re} s < \frac{1}{2}$) so that we can evaluate this last integral as an infinite sum. Using (4.6) again, inserting the definition (5.14), and interchanging summation and integration, we obtain

$$\int_0^\infty t^{2-s}\theta^{a/4n,b/4n}(t)\frac{dt}{t} = \pi^{s-2}\Gamma(2-s)\sum_{m\in\mathbb{Z}^2} m\cdot w\, e^{(2\pi i/4n)m\cdot(a,b)}|m|^{-2(2-s)}.$$

Thus, for $\operatorname{Re} 2 - s > \frac{3}{2}$, the right side of (5.17) is equal to

$$-i(4n)^{1-2s}\pi^{s-2}\Gamma(2-s)\frac{1}{4}\sum_{m\in\mathbb{Z}^2}\frac{m\cdot w}{|m|^{2(2-s)}}S_m \tag{5.18}$$

where for $m\in\mathbb{Z}^2$

$$S_m \underset{\text{def}}{=} \sum_{0\le a,b<4n} \chi_n'(a+bi)e^{(2\pi i/4n)m\cdot(a,b)}. \tag{5.19}$$

Lemma. *If $m_1 + m_2 i$ is not in the ideal $(1+i)$, then $S_m = 0$; whereas if $m_1 + m_2 i = (1+i)x$ with $x\in\mathbb{Z}[i]$, then $S_m = 2\chi_n'(x)g(\chi_n')$, where $g(\chi_n')$ is the Gauss sum defined in Proposition* 4 *of* §II.3 (*see* (3.9)).

Before proving the lemma, we show how the functional equation in the theorem follows immediately from it. Namely, if we make the substitution $m\cdot w = m_1 + m_2 i = (1+i)x$ in the sum in (5.18), the lemma gives us

$$\begin{aligned}\sum_{m\in\mathbb{Z}^2}\frac{m\cdot w}{|m|^{2(2-s)}}S_m &= \sum_{x\in\mathbb{Z}[i]}\frac{2(1+i)x}{|(1+i)x|^{2(2-s)}}\chi_n'(x)g(\chi_n') \\ &= (1+i)2^{s-1}\left(\frac{-2}{n}\right)(2+2i)n\sum_{x\in\mathbb{Z}[i]}\tilde{\chi}_n(x)(\mathbb{N}x)^{-(2-s)}\end{aligned}$$

by Proposition 4. But this last sum is $4L(E_n, 2-s)$, by (5.8). Bringing this all together, we conclude that for $\operatorname{Re} 2 - s > \frac{3}{2}$ the right side of (5.17) is equal to

$$\begin{aligned}&-i(4n)^{1-2s}\pi^{s-2}\Gamma(2-s)(1+i)2^{s-1}\left(\frac{-2}{n}\right)(2+2i)nL(E_n, 2-s) \\ &\quad = \left(\frac{-2}{n}\right)\pi^{s-2}\Gamma(2-s)(8n^2)^{1-s}L(E_n, 2-s).\end{aligned} \tag{5.20}$$

On the other hand, if we bring the term $(\sqrt{N}/2)^s$ over to the right in the functional equation (5.11)–(5.12) in the theorem, we find that what we want

to prove is:

$$\pi^{-s}\Gamma(s)L(E_n, s) = \left(\frac{-2}{n}\right)(\sqrt{N}/2)^{-s}(2\pi)^{s-2}(\sqrt{N})^{2-s}\Gamma(2-s)L(E_n, 2-s)$$
$$= \left(\frac{-2}{n}\right)(N/4)^{1-s}\pi^{s-2}\Gamma(2-s)L(E_n, 2-s).$$

And this is precisely (5.20).

Thus, to finish the proof of the theorem for odd n, it remains to prove the lemma.

Proof of Lemma. First suppose that $m_1 + m_2 i$ is not divisible by $1 + i$. This is equivalent to saying that m_1 and m_2 have opposite parity, i.e., their sum is odd. Now as a, b range from 0 to $4n$, the Gaussian integer $a + bi$ runs through each residue class modulo $(2 + 2i)n$ exactly twice. Each time gives the same value of $\chi_n'(a + bi)$, since $\chi_n'(a + bi)$ depends only on what $a + bi$ is modulo $n' = (2 + 2i)n$. But meanwhile, the exponential terms in the two summands have opposite sign, causing the two summands to cancel. To see this, we observe that if $a_1 + b_1 i$ and $a_2 + b_2 i$ are the two Gaussian integers in different residue classes modulo $4n$ but the same residue class modulo $(2 + 2i)n$, then $a_1 + b_1 i - (a_2 + b_2 i) \equiv (2 + 2i)n \pmod{4n}$, and so

$$e^{(2\pi i/4n)m\cdot((a_1,b_1)-(a_2,b_2))} = e^{(2\pi i/4n)m\cdot(2n,2n)} = e^{\pi i(m_1+m_2)} = -1.$$

This proves the first part of the lemma.

Now suppose that $m_1 + m_2 i = (1 + i)x$. Note that $m \cdot (a, b) = m_1 a + m_2 b = \mathrm{Re}((m_1 - m_2 i)(a + bi)) = \mathrm{Re}((1 - i)\bar{x}(a + bi))$. Hence, the exponential term in the summand in S_m is $\psi(\bar{x}(a + bi))$, where

$$\psi(x) \underset{\mathrm{def}}{=} e^{2\pi i\,\mathrm{Re}(x/n')}$$

(with $n' = (2 + 2i)n$). Since χ_n' is a primitive character modulo $(2 + 2i)n$ (see Proposition 3), we can apply Problem 9(a)–(b) of §II.2. Since the summation in (5.19) goes through each residue class modulo $(2 + 2i)n$ twice, we have

$$S_m = 2 \sum_{a+bi\in\mathbb{Z}[i]/(2+2i)n} \chi_n'(a + bi)\psi(\bar{x}(a + bi))$$
$$= 2\bar{\chi}_n'(\bar{x})g(\chi_n') = 2\chi_n'(x)g(\chi_n').$$

This proves the lemma, and hence the theorem (except for some slight modifications in the case of even n, which will be left as an exercise). □

In the problems we shall outline a proof of the analogous theorem in the case of an elliptic curve, namely $y^2 = x^3 + 16$, which has complex multiplication by another quadratic imaginary integer ring, namely $\mathbb{Z}[\omega]$, where $\omega = -\frac{1}{2} + \frac{1}{2}i\sqrt{3}$. There is one additional feature which is needed because we end up summing not over $\mathbb{Z}[i]$, which can be thought of as $\mathbb{Z}^2$, but rather over a lattice which is the image of $\mathbb{Z}^2$ under a certain 2×2-matrix.

So we have to apply Poisson summation to a function much like the function in this section, but involving this matrix.

We conclude this section by mentioning two references for a more general treatment of the theory of which we have only treated a few special cases. First, in C. L. Siegel's Tata notes [Siegel 1961] (see especially pp. 60–72) one finds L-functions whose summand has the form

$$e^{2\pi i m\cdot u}\frac{P(m+v)}{(Q[m+v])^{s+g/2}},$$

where $m\in\mathbb{Z}^n$, u, $v\in\mathbb{R}^n$, Q is the matrix of a positive definite quadratic form, and P is a "spherical polynomial with respect to Q of degree g". The case we needed for $L(E_n, s)$ was: $n=2$, $Q=\left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$, $P(x_1, x_2)=x_1+ix_2$, $g=1$. In Problem 8 below we have the case $Q=\left(\begin{smallmatrix}1&1/2\\1/2&1\end{smallmatrix}\right)$, $P(x_1, x_2)=(\omega+1/2)x_1+(\omega/2+1)x_2$ (where $\omega=-1/2+i\sqrt{3}/2$), $g=1$.

In [Lang 1970, Chapters XIII and XIV], two approaches are given to this topic. In Ch. XIII, the approach we have used (originally due to Hecke) is applied to obtain the functional equation for the Dedekind zeta-function of an arbitrary number field. This is a generalization of Problems 2 and 6 below. However, the case of more general Hecke L-series is not included in that chapter. A quite different approach due to J. Tate—using Fourier analysis on p-adic fields—is given in Ch. XIV of Lang's book.

Problems

1. Finish the proof of the theorem for n even.

2. (a) Find a functional equation for $\theta(t)\underset{\text{def}}{=}\sum_{m\in\mathbb{Z}^2}e^{-\pi t|m|^2}$, $t>0$.
   (b) The Dedekind zeta-function of a number field K is defined as follows:

   $$\zeta_K(s)\underset{\text{def}}{=}\sum(\mathbb{N}I)^{-s},$$

   where the sum is over all nonzero ideals I of the ring of integers of K. This sum converges for $\operatorname{Re} s>1$ (see [Borevich and Shafarevich 1966, Ch. 5, §1]). Let $K=\mathbb{Q}(i)$. Prove that $\zeta_K(s)$ is an entire function except for a simple pole at $s=1$ with residue $\pi/4$, and find a functional equation relating $\zeta_K(s)$ to $\zeta_K(1-s)$.

3. For u and v in $\mathbb{R}^2$, let

   $$\theta_u^v(t)\underset{\text{def}}{=}\sum_{m\in\mathbb{Z}^2}e^{2\pi i m\cdot v}e^{-\pi t|m+u|^2},\qquad t>0.$$

   Find a functional equation relating $\theta_u^v(t)$ to $\theta_{-v}^u(\frac{1}{t})$.

4. (a) In the situation of Proposition 11, express the Fourier transform of $(w\cdot\frac{\partial}{\partial x})^k f(x)$ in terms of $\hat{f}(y)$ for any nonnegative integer k.
   (b) Suppose that k is a nonnegative integer, $u\in\mathbb{R}^2$ is fixed with $u\notin\mathbb{Z}^2$, $t>0$ is fixed, and $w=(1, i)\in\mathbb{C}^2$. What is the Fourier transform of

   $$g(x)=((x+u)\cdot w)^k e^{-\pi t|x+u|^2}?$$

   (c) With k, u, t, w as in part (b), define:

$$\theta_{u,k}(t) = \sum_{m \in \mathbb{Z}^2} ((m+u)\cdot w)^k e^{-\pi t|m+u|^2};$$

$$\theta^{u,k}(t) = \sum_{m \in \mathbb{Z}^2} (m \cdot w)^k e^{2\pi i m \cdot u} e^{-\pi t|m|^2}.$$

Find a functional equation relating $\theta_{u,k}(t)$ to $\theta^{u,k}(\frac{1}{t})$.

(d) Suppose that I is a fixed ideal of $\mathbb{Z}[i]$, and χ: $(\mathbb{Z}[i]/I)^* \to \mathbb{C}^*$ is a nontrivial character. Outline (without computing the details) how one would prove that the function $\Sigma_{x \in \mathbb{Z}[i]} x^k \chi(x)(\mathbb{N}x)^{-s}$ extends to an entire function of s (if χ were trivial, there would be a simple pole at $s = 1$), and satisfies a functional equation relating its value at s to its value at $k + 1 - s$.

(e) Explain why any Hecke L-series for $\mathbb{Z}[i]$ is essentially of the form in part (d).

5. Let f: $\mathbb{R}^n \to \mathbb{C}$, $f \in \mathscr{S}$.

(a) For $M \in GL_n(\mathbb{R})$, let M^t denote the transpose matrix, and let $M^* = (M^{-1})^t$. Let $g(x) = f(Mx)$. Express $\hat{g}(y)$ in terms of $\hat{f}(y)$.

(b) Let L be a lattice in $\mathbb{R}^n$; equivalently, L is of the form $L = M\mathbb{Z}^n$, where $M \in GL_n(\mathbb{R})$, i.e., L is obtained by applying some matrix M to the elements in the standard basis lattice $\mathbb{Z}^n$. Let L' be the "dual lattice" defined by: $L' = \{y \in \mathbb{R}^n | x \cdot y \in \mathbb{Z}$ for all $x \in L\}$. First show that L' is a lattice. Then prove a functional equation relating $\Sigma_{x \in L} f(x)$ to $\Sigma_{y \in L'} \hat{f}(y)$.

6. Let $\omega = -\frac{1}{2} + \frac{1}{2}\sqrt{3}$.

(a) Let $M = \begin{pmatrix} 1 & -1/2 \\ 0 & \sqrt{3/2} \end{pmatrix}$. Show that the lattice $L = \mathbb{Z}[\omega] \subset \mathbb{C} = \mathbb{R}^2$ is $L = M\mathbb{Z}^2$. What are M^* and L'? Show that L', considered as a lattice in $\mathbb{C}$, is twice the "different" of $\mathbb{Z}[\omega]$, which is defined as $\{x \in \mathbb{Q}(\omega) | \text{Tr}(xy) \in \mathbb{Z}$ for all $y \in \mathbb{Z}[\omega]\}$. Here $\text{Tr}\, x = x + \bar{x} = 2\,\text{Re}\, x$.

(b) Prove that

$$\sum_{x \in \mathbb{Z}[\omega]} e^{-\pi t|x|^2} = \frac{2}{t\sqrt{3}} \sum_{x \in \mathbb{Z}[\omega]} e^{-(4\pi/3t)|x|^2}.$$

(c) Let $K = \mathbb{Q}(\omega)$. Prove that $\zeta_K(s)$ extends meromorphically onto the complex s-plane with its only pole a simple pole at $s = 1$. Find the residue at the pole, and derive a functional equation for $\zeta_K(s)$.

(d) With K as in part (c), prove the identity: $\zeta_K(s) = \zeta(s)L(\chi, s)$, where χ is the nontrivial character of $(\mathbb{Z}/3\mathbb{Z})^*$.

(e) Use part (d), the theorem in §II.4, and Problem 5(e) of §II.4 to give another proof of the functional equation in part (c).

7. Let E be the elliptic curve $y^2 = x^3 + 16$. Let $\omega = -\frac{1}{2} + \frac{i}{2}\sqrt{3}$.

(a) Show that the reduction mod p of $y^2 = x^3 + 16$ is an elliptic curve over $\mathbb{F}_p$ if and only if $p \neq 2, 3$. For such p, recall that the Euler factor at p is $(1 - 2a_p p^{-s} + p^{1-2s})^{-1}$, where $(1 - 2a_p T + pT^2)$ is the numerator of $Z(E/\mathbb{F}_p; T)$. Show that for $p \neq 2, 3$ this factor is

$$\prod_P (1 - \alpha_P^{\deg P} \mathbb{N}P^{-s})^{-1},$$

where the product is over all (one or two) prime ideals P of $\mathbb{Z}[\omega]$ dividing (p) and $\alpha_P^{\deg P}$ is the unique generator of P such that $\alpha_P^{\deg P} \equiv 1 \pmod 3$.

(b) When $p = 3$, define the Euler factor at 3 to be simply 1, as we did for E_n: $y^2 = x^3 - n^2 x$ when $p | 2n$. When $p = 2$, at first glance it looks like we again have

$y^2 = x^3$, and so we might be tempted to take 1 for the Euler factor at 2 as well. However, this is wrong. When one defines $L(E, s)$, if there exists a $\mathbb{Q}$-linear change of variables that takes our equation E: $y^2 = x^3 + 16$ in $\mathbb{P}^2_{\mathbb{Q}}$ to a curve C whose reduction mod p is smooth, then we are obliged to say that E has good reduction at p and to form the corresponding Euler factor from the zeta-function of C over $\mathbb{F}_p$. In Problem 22 of §II.2, we saw that $y^2 = x^3 + 16$ is equivalent over $\mathbb{Q}$ to $y^2 + y = x^3$, a smooth curve over $\mathbb{F}_2$ whose zeta-function we computed. Show that the Euler factor at $p = 2$ is given by the same formula as in part (a).

(c) For $x \in \mathbb{Z}[\omega]$ prime to 3, let $\chi(x) = (-\omega)^j$ be the unique sixth root of 1 such that $x\chi(x) \equiv 1 \pmod 3$. Let $\chi(x) = 0$ for x in the ideal $(\sqrt{-3})$. Show that

$$L(E, s) = \frac{1}{6} \sum_{x \in \mathbb{Z}[\omega]} \frac{\chi(x)x}{(\mathbb{N}x)^s}.$$

(d) Let

$$\psi(x) \underset{\text{def}}{=} e^{(2\pi i/3)\operatorname{Tr}(x/i\sqrt{3})},$$

where $\operatorname{Tr} x = x + \bar{x} = 2 \operatorname{Re} x$. Verify that $\psi(x)$ is an additive character of $\mathbb{Z}[\omega]/3$ satisfying the condition in Problem 9 of §II.2 (i.e., that it is nontrivial on any larger ideal than (3)). Find the value of $g(\chi, \psi) = \sum_{x \in \mathbb{Z}[\omega]/3} \chi(x)\psi(x)$, where χ is as in part (b).

8. (a) Let $w = (1, \omega)$, $u \in \mathbb{R}^2 - \mathbb{Z}^2$, and $t > 0$ be fixed. What is the Fourier transform of $g(x) = (x + u) \cdot w\, e^{-\pi t |(x+u) \cdot w|^2}$?

(b) Let $\theta_u(t) = \sum_{m \in \mathbb{Z}^2} g(m)$, where $g(x)$ is as in part (a). Find the Mellin transform $\phi(s)$ of

$$\sum_{\substack{u=(a/3,b/3)\\ 0 \le a,b < 3}} \chi(a + b\omega)\theta_u(t).$$

(c) Prove that $L(E, s)$ is an entire function, where E: $y^2 = x^3 + 16$ is the elliptic curve in Problem 7.

(d) Prove the functional equation $\Lambda(s) = \Lambda(2 - s)$ with

$$\Lambda(s) \underset{\text{def}}{=} \left(\frac{\sqrt{27}}{2\pi}\right)^s \Gamma(s) L(E, s).$$

§6. The critical value

The value at $s = 1$ of the Hasse–Weil L-function $L(E, s)$ of an elliptic curve E is called the "critical value". When we have a functional equation relating $L(E, s)$ to $L(E, 2 - s)$, the point $s = 1$ is the "center" of the functional equation, in the sense that it is the fixed point of the correspondence $s \leftrightarrow 2 - s$. The importance of this critical value comes from the following famous conjecture.

Conjecture (B. J. Birch and H. P. F. Swinnerton-Dyer). $L(E, 1) = 0$ *if and only if* E *has infinitely many rational points.*

In this conjecture E is any elliptic curve defined over $\mathbb{Q}$. In the general case it has not even been proved that it makes sense to speak of $L(E, 1)$,

because no one has been able to prove analytic continuation of $L(E, s)$ to the left of the line Re $s = \frac{3}{2}$. However, analytic continuation and a functional equation have been proved for any elliptic curve with complex multiplication (see Problem 8 of §I.8), of which our E_n are special cases, and for a broader class of elliptic curves with a so-called "Weil parametrization" by modular curves. (It has been conjectured by Weil and Taniyama that the latter class actually consists of all elliptic curves defined over $\mathbb{Q}$.)

We shall call the above conjecture the "weak Birch–Swinnerton-Dyer conjecture", because Birch and Swinnerton-Dyer made a much more detailed conjecture about the behavior of $L(E, s)$ at $s = 1$. Namely, they conjectured that the *order* of zero is equal to the *rank* r of the group of rational points on E (see the beginning of §I.9). Moreover, they gave a conjectural description of the coefficient of the first nonvanishing term in the Taylor expansion at $s = 1$ in terms of various subtle arithmetic properties of E. For a more detailed discussion of the conjecture of Birch and Swinnerton-Dyer, see [Birch 1963], [Birch and Swinnerton-Dyer 1963, 1965], [Cassels 1966], [Swinnerton-Dyer 1967], [Tate 1974].

There is a simple heuristic argument—far from a proof—which shows why the weak Birch–Swinnerton-Dyer conjecture might be true. Let us pretend that the Euler product for $L(E, s)$ (see (5.1) for the case $E = E_n$) is a convergent infinite product when $s = 1$ (which it probably isn't). In that case we would have:

$$L(E, 1) = \prod_p \frac{1}{1 - 2a_{E,p}p^{-s} + p^{1-2s}}\bigg|_{s=1} = \prod_p \frac{p}{p + 1 - 2a_{E,p}} = \prod_p \frac{p}{N_p},$$

where $N_p = N_{1,p} = p + 1 - 2a_{E,p}$ is the number of $\mathbb{F}_p$-points on the elliptic curve E considered modulo p. Now as p varies, the N_p "straddle" p at a distance bounded by $2\sqrt{p}$. This is because $2a_{E,p} = \alpha_p + \bar{\alpha}_p$, and the reciprocal roots α_p have absolute value $\sqrt{p}$ (see (2.7) for $E = E_n$, and the discussion of the Weil conjectures in §1 for the general case). Thus, roughly speaking, $N_p \approx p \pm \sqrt{p}$. If N_p spent an equal time on both sides of p as p varies, one could expect the infinite product of the p/N_p to converge to a nonzero limit. (See Problem 1 below.) If, on the other hand, the N_p had a tendency to be on the large side: $N_p \approx p + \sqrt{p}$, then we would obtain $L(E, 1) \approx \Pi_p p/(p + \sqrt{p}) = \Pi_p 1/(1 + p^{-1/2}) = 0$.

To conclude this heuristic argument, we point out that, if there are infinitely many rational points, one would expect that by reducing them modulo p (as in the proof of Proposition 17 in §I.9) we would obtain a large guaranteed contribution to N_p for all p, thereby ensuring this lopsided behavior $N_p \approx p + \sqrt{p}$. On the other hand, if there are only finitely many rational points, then their contribution to N_p would be negligible for large p, so that N_p would have the "random" behavior $N_p \approx p \pm \sqrt{p}$. Needless to say, this heuristic argument is not of much help in trying to prove the weak Birch–Swinnerton-Dyer conjecture.

However, there is a remarkable result due to D. Goldfeld which states the following [Goldfeld 1982]: *if* the infinite product $\prod_p (N_p/p)$ does converge to

a nonzero limit, then (a) that limit is $\sqrt{2}/L(E, 1)$, i.e., it is equal not to the reciprocal of the critical value (as in the above heuristic argument) but rather to the reciprocal of the critical value *multiplied by* $\sqrt{2}$; and (b) the analogue of the Generalized Riemann Hypothesis holds for the Hasse–Weil L-function $L(E, S)$. Because of part (b), no one expects to be able to prove convergence of $\prod_p (N_p/p)$ any time soon (even for a single specific case).

The really convincing evidence to support the conjecture of Birch and Swinnerton-Dyer comes not from the above heuristics or computational examples, but rather from a series of dramatic partial results in the direction of the conjecture. The first breakthrough came in 1977, when Coates and Wiles proved that for a large class of elliptic curves, an infinite number of rational points implies that $L(E, 1) = 0$. (Other partial results will be discussed briefly at the end of the book.)

Recall from Problem 8 of §I.8 that an elliptic curve is said to have complex multiplication if its lattice is taken to itself under multiplication by some complex numbers other than integers.

Theorem (J. Coates and A. Wiles). *Let E be an elliptic curve defined over $\mathbb{Q}$ and having complex multiplication. If E has infinitely many $\mathbb{Q}$-points, then $L(E, 1) = 0$.*

The proof of this theorem is rather difficult (see [Coates and Wiles 1977]), and it will not be given here. (The original proof further assumed that the quadratic imaginary field of complex multiplication has class number 1, but this turned out not to be necessary.)

Since our curves E_n have complex multiplication, the Coates–Wiles theorem applies, and, in view of Proposition 18 of Chapter I, tells us that if $L(E_n, 1) \neq 0$, then n is not a congruent number. Conversely, if we allow ourselves the weak Birch–Swinnerton-Dyer conjecture, then it follows that $L(E_n, 1) = 0$ implies that n is a congruent number.

There is one situation in which it is easy to know that $L(E_n, 1) = 0$. Recall that the "root number"—the plus or minus sign in the functional equation for $L(E_n, s)$—is equal to $(\frac{-2}{n})$ for n odd, and $(\frac{-1}{n_0})$ for $n = 2n_0$ even (see the theorem in §5).

Proposition 12. *If $n \equiv 5$, 6 or 7 (mod 8), and if the weak Birch–Swinnerton-Dyer conjecture holds for E_n, then n is a congruent number.*

Proof. According to the theorem in §5, if $n \equiv 5, 6, 7 \pmod 8$, then $\Lambda(s) = -\Lambda(2 - s)$, where $\Lambda(s)$ is given by (5.11). Substituting $s = 1$, we conclude that $\Lambda(1) = -\Lambda(1)$, i.e., $\Lambda(1) = 0$. But by (5.11), $\Lambda(1)$ differs from $L(E_n, 1)$ by a nonzero factor (namely, $\sqrt{N}/2\pi$). Thus, $L(E_n, 1) = 0$, and the weak Birch–Swinnerton-Dyer conjecture then tells us that E_n has infinitely many $\mathbb{Q}$-points, i.e., by Proposition 18 of Chapter I, n is a congruent number. □

In certain cases, the claim that all $n \equiv 5, 6, 7 \pmod 8$ are congruent numbers has been verified without assuming the weak Birch–Swinnerton-Dyer

conjecture. A method due to Heegner (see [Birch 1975]) for constructing points on E_n enables one to prove this claim for n equal to a prime or twice a prime. That is, if n is a prime congruent to 5 or 7 modulo 8, or twice a prime congruent to 3 mod 4, then n is known to be a congruent number.

It is interesting to note that even in the cases when Heegner's method ensures us that n is a congruent number, the method does not give us an effective algorithm for constructing a nontrivial rational point on E_n, or equivalently, finding a right triangle with rational sides and area n.

Somewhat later, Gross and Zagier were able to improve greatly upon Heegner's method. As a special case of their results, they showed that for $n \equiv 5, 6, 7 \pmod 8$ the elliptic curve E_n has infinitely many rational points provided that $L(E_n, s)$ has only a simple zero at $s = 1$, i.e., $L'(E_n, 1) \neq 0$. This result represents substantial progress in making Proposition 12 unconditional. Moreover, their method is constructive, i.e., it gives you a rational point on the curve (equivalently, a right triangle with area n) when $L'(E_n, 1) \neq 0$. See [Gross and Zagier 1983 and 1986] and [Coates 1986].

In the cases when the root number is $+1$, we cannot be sure in advance whether $L(E_n, 1)$ is zero or nonzero. So in those cases it is useful to have an efficient algorithm for computing $L(E_n, 1)$, at least to enough accuracy to know for certain that it is nonzero. (It is harder to be sure of ourselves in cases when the critical value seems to be zero.) We cannot use the series (5.3) or (5.8) to evaluate $L(E_n, 1)$, since they only converge when $\operatorname{Re} s > \frac{3}{2}$.

So we now turn our attention to finding a rapidly convergent expression for $L(E_n, 1)$.

Let us return to the functional equation for $L(E_n, s)$, and give a slightly different, more efficient proof. Recall that

$$L(E_n, s) = \frac{1}{4} \sum_{x \in \mathbb{Z}[i]} \tilde{\chi}_n(x)(\mathbb{N}x)^{-s}$$

with $\tilde{\chi}_n(x) = x\chi_n'(x)$, where χ_n' was defined in (3.3)–(3.4). Suppose we ask the question, "What function $F(E_n, t)$ has $\pi^{-s}\Gamma(s)L(E_n, s)$ as its Mellin transform?" By our usual method using (4.6), we see that the answer is

$$F(E_n, t) \underset{\text{def}}{=} \frac{1}{4} \sum_{x \in \mathbb{Z}[i]} \tilde{\chi}_n(x) e^{-\pi t |x|^2}. \tag{6.1}$$

We now proceed to find a functional equation for $F(E_n, t)$, which will then immediately give us once again our functional equation for the Mellin transform $L(E_n, s)$. The only difference with our earlier derivation is whether we take the character sum before or after applying the functional equation (compare with the two derivations in Problems 3 and 5 of §II.4 and in Problems 4 and 6 for the functional equation for a Dirichlet L-series).

Recall that $\chi_n'(x)$ is a primitive character of $(\mathbb{Z}[i]/n')^*$, where $n' = (2 + 2i)n$ for n odd, $n' = 2n$ for n even. Let $a + bi$ run through some set of coset representatives of $\mathbb{Z}[i]$ modulo n', and for each $a + bi$ define a corresponding pair (u_1, u_2) of rational numbers by: $u_1 + u_2 i = (a + bi)/n'$. Replacing x by $a + bi + n'(m \cdot (1, i))$ for $m \in \mathbb{Z}^2$, and setting

$$N' = |n'|^2 = \frac{N}{4} \tag{6.2}$$

(see (5.10)), we can rewrite $F(E_n, t)$ as follows:

$$F(E_n, t) = \frac{n'}{4} \sum_{a+bi \in \mathbb{Z}[i]/n'} \chi'_n(a+bi) \sum_{m \in \mathbb{Z}^2} (m+u)\cdot(1, i)e^{-\pi N' t|m+u|^2}.$$

If we replace t by $\frac{1}{N't}$, the summand in the inner sum becomes $\theta_u(\frac{1}{t})$ in the notation of (5.13). We then use the functional equation (5.16) for $\theta_u(\frac{1}{t})$. As a result we obtain:

$$\begin{aligned} F\left(E_n, \frac{1}{N't}\right) &= \frac{n'}{4} \sum_{a+bi} \chi'_n(a+bi)(-it^2) \sum_m m\cdot(1, i)e^{2\pi i m\cdot u}e^{-\pi t|m|^2} \\ &= -\frac{i}{4}t^2 n' \sum_m m\cdot(1, i)e^{-\pi t|m|^2} \sum_{a+bi} \chi'_n(a+bi)e^{2\pi i m\cdot u}. \end{aligned}$$

Now $m\cdot u = \mathrm{Re}((m_1 - m_2 i)(u_1 + u_2 i)) = \mathrm{Re}((m_1 - m_2 i)(a+bi)/n')$. We now use Problem 2 of §II.2 to rewrite the last inner sum as

$$\bar{\chi}'_n(m_1 - m_2 i) \sum_{a+bi} \chi'_n(a+bi)e^{2\pi i \,\mathrm{Re}((a+bi)/n')}.$$

But $\bar{\chi}'_n(m_1 - m_2 i) = \chi'_n(m_1 + m_2 i)$, and the sum here is the Gauss sum $g(\chi'_n)$ evaluated in Proposition 4 (see (3.9)). We finally obtain:

$$F\left(E_n, \frac{1}{N't}\right) = -\frac{i}{4}t^2 n'(\varepsilon n') \sum_{m \in \mathbb{Z}^2} \bar{\chi}_n(m_1 + m_2 i)e^{-\pi t|m|^2},$$

where the ε is $(\frac{-2}{n})$ for n odd, $i(\frac{-1}{n_0})$ for $n = 2n_0$ even. Replacing $m_1 + m_2 i$ by $x \in \mathbb{Z}[i]$, we see that the sum is precisely $4F(E_n, t)$. Thus, we have

$$F\left(E_n, \frac{1}{N't}\right) = \begin{cases} \left(\dfrac{-2}{n}\right) N't^2 F(E_n, t), & n \text{ odd}; \\ \left(\dfrac{-1}{n_0}\right) N't^2 F(E_n, t), & n = 2n_0 \text{ even}. \end{cases} \tag{6.3}$$

We can easily derive the functional equation for $L(E_n, s)$ from (6.3). We shall write $\pm$ to denote $(\frac{-2}{n})$ for n odd, $(\frac{-1}{n_0})$ for $n = 2n_0$ even. We have

$$\pi^{-s}\Gamma(s)L(E_n, s) = \int_0^\infty t^s F(E_n, t)\frac{dt}{t} = \pm\frac{1}{N'}\int_0^\infty t^{s-2}F\left(E_n, \frac{1}{N't}\right)\frac{dt}{t}$$

by (6.3). Making the change of variables $u = \frac{1}{N't}$, we obtain:

$$\begin{aligned} \pi^{-s}\Gamma(s)L(E_n, s) &= \pm N'^{1-s}\int_0^\infty u^{2-s}F(E_n, u)\frac{du}{u} \\ &= \pm N'^{1-s}\pi^{s-2}\Gamma(2-s)L(E_n, 2-s). \end{aligned}$$

Finally, replacing N'^{1-s} by $(\sqrt{N}/2)^{2-2s}$ and multiplying both sides by $(\sqrt{N}/2)^s$, we obtain the functional equation of (5.11)–(5.12).

We now use our function $F(E_n, t)$ in the case when the root number is $+1$, i.e., $n \equiv 1, 2, 3 \pmod 8$, in order to find a convenient expression for $L(E_n, 1)$. Thus, suppose that

$$F\left(E_n, \frac{1}{N't}\right) = N't^2 F(E_n, t). \tag{6.4}$$

We use this functional equation to break up the Mellin transform of $F(E_n, t)$ into two integrals from $\frac{1}{\sqrt{N'}}$ to ∞. The point $\frac{1}{\sqrt{N'}}$ is the "center" of the functional equation, i.e., the fixed point of the correspondence $t \leftrightarrow \frac{1}{N't}$.

We have:

$$\pi^{-s}\Gamma(s)L(E_n, s) = \int_0^\infty t^s F(E_n, t)\frac{dt}{t} = \int_{1/\sqrt{N'}}^\infty + \int_0^{1/\sqrt{N'}} t^s F(E_n, t)\frac{dt}{t}.$$

In the second integral we replace t by $\frac{1}{N't}$, and then use (6.4) to write $F(E_n, \frac{1}{N't})$ in terms of $F(E_n, t)$. The result is:

$$\pi^{-s}\Gamma(s)L(E_n, s) = \int_{1/\sqrt{N'}}^\infty (t^s F(E_n, t) + N'^{1-s}t^{2-s}F(E_n, t))\frac{dt}{t}.$$

Now set $s = 1$. Multiplying both sides by π, we immediately obtain:

$$L(E_n, 1) = 2\pi \int_{1/\sqrt{N'}}^\infty F(E_n, t)dt. \tag{6.5}$$

Recall that the Hasse–Weil L-function can be written as a Dirichlet series

$$L(E_n, s) = \sum_{m=1}^\infty b_{m,n} m^{-s}, \quad \text{where} \quad b_{m,n} = \frac{1}{4} \sum_{\substack{x \in \mathbb{Z}[i] \\ \mathbb{N}x = m}} \tilde{\chi}_n(x). \tag{6.6}$$

Comparing with the definition (6.1) of $F(E_n, t)$, we see that

$$F(E_n, t) = \sum_{m=1}^\infty b_{m,n} e^{-\pi t m}. \tag{6.7}$$

We can now substitute the series (6.7) into (6.5) and integrate term by term. (Notice that the procedure below will work only because we have a *positive* lower limit of integration in (6.5); if we tried directly to use the Mellin transform, in which the lower limit of integration is 0, we would not have convergence.) Using the formula $\int_a^\infty e^{-ct}dt = \frac{1}{c}e^{-ac}$ with $a = N'^{-1/2}$, $c = \pi m$, we immediately obtain the following rapidly convergent infinite series for $L(E_n, 1)$.

Proposition 13. *The critical value of the Hasse–Weil L-function of the elliptic curve $E_n\colon y^2 = x^3 - n^2 x$ for squarefree $n \equiv 1, 2, 3 \pmod 8$ is given by:*

$$L(E_n, 1) = 2\sum_{m=1}^\infty \frac{b_{m,n}}{m} e^{-\pi m/\sqrt{N'}}, \quad \textit{where} \quad \sqrt{N'} = \begin{cases} 2n\sqrt{2}, & n \textit{ odd;} \\ 2n, & n \textit{ even.} \end{cases} \tag{6.8}$$

Here the coefficients $b_{m,n}$ are the Dirichlet series coefficients obtained by

expanding

$$L(E_n, s) = \prod_{p \nmid 2n} (1 - 2a_{E_n,p}p^{-s} + p^{1-2s})^{-1} = \sum_{m=1}^{\infty} b_{m,n} m^{-s}.$$

In addition, the absolute value of the coefficient $b_{m,n}$ is bounded by $\sigma_0(m)\sqrt{m}$, where $\sigma_0(m)$ denotes the number of divisors of m.

PROOF. We have already proved all except for the bound on $b_{m,n}$. If we write the Euler factor in the form $(1 - \alpha_p p^{-s})^{-1}(1 - \bar{\alpha}_p p^{-s})^{-1}$, expand each factor in a geometric series, and collect coefficients of p^{-es} for each positive integer e, we find that the coefficient of p^{-es} is $\alpha_p^e + \alpha_p^{e-1}\bar{\alpha}_p + \alpha_p^{e-2}\bar{\alpha}_p^2 + \cdots + \bar{\alpha}_p^e$. This means that, if m has prime factorization $m = p_1^{e_1} \cdots p_r^{e_r}$, then

$$b_{m,n} = \prod_{j=1}^{r} (\alpha_{p_j}^{e_j} + \alpha_{p_j}^{e_j-1}\bar{\alpha}_{p_j} + \cdots + \bar{\alpha}_{p_j}^{e_j}).$$

Since $|\alpha_p| = |\bar{\alpha}_p| = \sqrt{p}$ for all p, we immediately obtain the bound

$$|b_{m,n}| \leq \prod_{j=1}^{r} (e_j + 1)p_j^{e_j/2} = \sigma_0(m)\sqrt{m},$$

where we have used the easy fact from elementary number theory that $\sigma_0(m)$ is the product of the $(e_j + 1)$. This completes the proof. □

The bound for $b_{m,n}$ is useful in estimating the remainder after we compute the series in (6.8) out to the M-th place. In particular, if we find that the remainder is less than the value of that partial sum, we may conclude that $L(E_n, 1) \neq 0$.

As an example, we treat the case $n = 1$. The first few Dirichlet series coefficients b_m for $L(E_1, s)$ are given in (5.4). By (6.8), we have

$$\begin{aligned} L(E_1, 1) &\\ &= 2(e^{-\pi/2\sqrt{2}} - \tfrac{2}{5}e^{-5\pi/2\sqrt{2}} - \tfrac{1}{3}e^{-9\pi/2\sqrt{2}} + \tfrac{6}{13}e^{-13\pi/2\sqrt{2}} + \tfrac{2}{17}e^{-17\pi/2\sqrt{2}} + \cdots) \\ &= 0.6555143\ldots + R_{25}, \end{aligned}$$

where we have denoted $R_M = 2\sum_{m \geq M} \frac{b_m}{m} e^{-\pi m/2\sqrt{2}}$.

A very crude estimate for $\sigma_0(m)$ is $2\sqrt{m}$ (see the problems). Thus,

$$|R_M| \leq 4 \sum_{m \geq M} e^{-\pi m/2\sqrt{2}} = \frac{4}{1 - e^{-\pi/2\sqrt{2}}} e^{-\pi M/2\sqrt{2}}.$$

So R_{25} is bounded by 5.2×10^{-12}. Actually, the convergence is so fast that we only needed to evaluate the first term to show that $L(E_1, 1) \neq 0$:

$$L(E_1, 1) = 0.6586\ldots + R_5, \quad \text{with} \quad |R_5| \leq 0.023.$$

This computation, together with the Coates–Wiles theorem, then tells us that 1 is not a congruent number. In fact, this argument undoubtedly qualifies as the world's most roundabout proof of that fact, which was proved by

Fermat more than three hundred years ago. (See [Weil 1973, p. 270 of Vol. III of Collected Papers]; see also Problem 3 of §I.1.)

The next topic we take up is the systematic study of functions such as theta-series which have certain types of functional equations under $t \mapsto \frac{1}{t}$ and similar changes of variable. Such functions are called "modular forms". Actually, modular forms are functions of the form $\Sigma\, b_m e^{2\pi i z}$ rather than $\Sigma\, b_m e^{-\pi t}$, but the simple substitution $t = -2iz$ will transform our theta-series from this chapter into what turns out to be modular forms.

In studying modular forms, we will at the same time be approaching elliptic curves from another perspective. But these two aspects of elliptic curves—the congruence zeta-function and Hasse–Weil L-series, and the theory of modular forms—have combined in recent years to form a richly interlocking picture.

Problems

1. In the heuristic argument for the weak Birch–Swinnerton-Dyer conjecture, make the following ridiculous assumptions: (i) $|2a_{E,p} - 1| = \sqrt{p}$; and (ii) $(2a_{E,p} - 1)/\sqrt{p} = \pm 1$ is "evenly" distributed, and happens to coincide with the value at p of a fixed quadratic Dirichlet character $\chi(p) = (\frac{p}{N})$ for some fixed N. (One of the reasons why these assumptions are ridiculous is that $2a_{E,p}$ is an integer.) Show that then $L(E, 1)$ is equal to the value $L(\chi, \frac{1}{2})$ of the Dirichlet L-function at the center of symmetry of its functional equation.

2. Prove that if the root number in the functional equation for $L(E_n, s)$ is 1, then $L(E_n, s)$ has either a nonzero value or else an even-order zero at $s = 1$; and if the root number is -1, then $L(E_n, s)$ has an odd-order zero at $s = 1$.

3. In the notation of Proposition 13 (here we abbreviate $b_m = b_{m,n}$, $a_p = a_{E_n,p}$), prove that:
   (a) $b_p = 2a_p$ if $p \nmid 2n$; $b_p = 0$ if $p | 2n$;
   (b) $b_{m_1 m_2} = b_{m_1} b_{m_2}$ if m_1 and m_2 are relatively prime;
   (c) $b_{p^{e+1}} = 2a_p b_{p^e} - p b_{p^{e-1}}$ for $e \geq 0$ (here take $b_{1/p} = 0$ when $e = 0$).

4. Prove that $\sigma_0(m) < 2\sqrt{m}$ for all m, and that $m^{-\varepsilon}\sigma_0(m) \to 0$ as $m \to \infty$ for any positive ε.

5. Compute $L(E_2, 1)$ and $L(E_3, 1)$ to about three decimal places of accuracy, verifying that they are nonzero.

6. Prove that $L(E_{10}, 1) \neq 0$.

7. Suppose you knew a lower bound c for the absolute value of all nonzero $L(E_n, 1)$, $n = 1, 2, 3, \ldots$ squarefree. (No such c is known.) For n very large, what is the order of magnitude of M such that you could determine from the first M terms in (6.8) whether or not $L(E_n, 1) = 0$?

8. (a) Write a flow chart for a computer program that evaluates $L(E_n, 1)$ through the M-th term of (6.8) and estimates the remainder.
   (b) If you have a computer handy, use part (a) to find $L(E_{41}, 1)$ to three decimal places.

CHAPTER III

Modular Forms

Our treatment of the introductory material will be similar to that in Serre's *A Course in Arithmetic* (Chapter VII), except that we shall bring in the "level" from the very beginning.

§1. $SL_2(\mathbb{Z})$ and its congruence subgroups

For any commutative ring R, the "general linear group" $GL_2(R)$ is defined to be the set of matrices $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ such that $\det g = ad - bc$ is in R^* (the multiplicative group of invertible elements of R). It is easy to see that $GL_2(R)$ is a group. The "special linear group" $SL_2(R)$ is defined to be the subgroup of $GL_2(R)$ consisting of matrices of determinant 1. In this section we shall be concerned with the cases $R = \mathbb{R}$ (the real numbers), $R = \mathbb{Z}$, $R = \mathbb{Z}/N\mathbb{Z}$ (for a positive integer N).

Let $\tilde{\mathbb{C}}$ denote $\mathbb{C} \cup \{\infty\}$ (i.e., the complex plane with a point at infinity, or equivalently the complex projective line $\mathbb{P}^1_{\mathbb{C}}$, also known as the "Riemann sphere"). Given an element $g = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in SL_2(\mathbb{R})$ and a point $z \in \mathbb{C}$, we define

$$gz \underset{\text{def}}{=} \frac{az+b}{cz+d}; \qquad g\infty \underset{\text{def}}{=} a/c = \lim_{z\to\infty} gz. \tag{1.1}$$

(Thus, $g(-d/c) = \infty$, and if $c = 0$, then $g\infty = \infty$.) These maps $z \mapsto gz$ are called "fractional linear transformations" of the Riemann sphere $\tilde{\mathbb{C}}$. It is easy to check that (1.1) defines a group action on the set $\tilde{\mathbb{C}}$, in other words: $g_1(g_2 z) = (g_1 g_2)z$ for all $g_1, g_2 \in SL_2(\mathbb{R})$, $z \in \tilde{\mathbb{C}}$.

Notice that for $g = -I = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right) \in SL_2(\mathbb{R})$, (1.1) gives the identity map.

But $\pm I$ are the only matrices which act trivially on $\tilde{\mathbb{C}}$, as we see by supposing that $z = (az + b)/(cz + d)$ for all z, i.e., $cz^2 + (d - a)z - b = 0$ for all z, which implies that $b = c = 0$ and $a = d$; but the only scalar matrices $\left(\begin{smallmatrix} a & 0 \\ 0 & a \end{smallmatrix}\right)$ of determinant 1 are $\pm I$. Thus, the quotient group $SL_2(\mathbb{R})/\pm I$, which is sometimes denoted $PSL_2(\mathbb{R})$, acts "faithfully" on $\mathbb{C}$, in other words, each element other than the identity acts nontrivially.

Let $H \subset \mathbb{C}$ denote the upper half-plane $H = \{z \in \mathbb{C} | \operatorname{Im} z > 0\}$. It is important to note that any $g \in SL_2(\mathbb{R})$ preserves H, i.e., $\operatorname{Im} z > 0$ implies $\operatorname{Im} gz > 0$. This is because

$$\operatorname{Im} gz = \operatorname{Im} \frac{az + b}{cz + d} = \operatorname{Im} \frac{(az + b)(c\bar{z} + d)}{|cz + d|^2} = |cz + d|^{-2} \operatorname{Im}(adz + bc\bar{z}).$$

But $\operatorname{Im}(adz + bc\bar{z}) = (ad - bc)\operatorname{Im} z = \operatorname{Im} z$, since $\det g = 1$. Thus,

$$\operatorname{Im} gz = |cz + d|^{-2} \operatorname{Im} z \quad \text{for} \quad g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{R}). \tag{1.2}$$

Thus, the group $SL_2(\mathbb{R})$ acts on the set H by the transformations (1.1).

The subgroup of $SL_2(\mathbb{R})$ consisting of matrices with integer entries is, by definition, $SL_2(\mathbb{Z})$. It is called the "full modular group", and is sometimes denoted Γ. We shall denote $\bar{\Gamma} \underset{\text{def}}{=} \Gamma/\pm I$. (Whenever we have a subgroup G of $SL_2(\mathbb{R})$, we shall let $\bar{G}$ denote $G/\pm I$ if G contains $-I$; if $-I \notin G$, we set $\bar{G} = G$.) This group $\bar{\Gamma} = SL_2(\mathbb{Z})/\pm I$ acts faithfully on H. It is one of the basic groups arising in number theory and other branches of mathematics.

Besides $\Gamma = SL_2(\mathbb{Z})$, certain of its subgroups have special significance. Let N be a positive integer. We define

$$\Gamma(N) \underset{\text{def}}{=} \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}) \middle| a \equiv d \equiv 1 \pmod{N},\ b \equiv c \equiv 0 \pmod{N} \right\}. \tag{1.3}$$

This is a subgroup of $\Gamma = SL_2(\mathbb{Z})$, actually a normal subgroup, because it is the kernel of the group homomorphism from Γ to $SL_2(\mathbb{Z}/N\mathbb{Z})$ obtained by reducing entries modulo N. In other words, $\Gamma(N)$ consists of 2×2 integer matrices of determinant 1 which are congruent modulo N to the identity matrix. Note that $\Gamma(1) = \Gamma$. We shall later see that $\Gamma(N)$ is analogous to the subsemigroup $1 + N\mathbb{Z} \subset \mathbb{Z}$ consisting of integers congruent to 1 modulo N. $\Gamma(N)$ is called the "principal congruence subgroup of level N".

Notice that $\bar{\Gamma}(2) = \Gamma(2)/\pm I$, whereas for $N > 2$ we have $\bar{\Gamma}(N) = \Gamma(N)$, because $-1 \not\equiv 1 \pmod{N}$, and so $-I$ is not in $\Gamma(N)$.

A subgroup of Γ (or of $\bar{\Gamma}$) is called a "congruence subgroup of level N" if it contains $\Gamma(N)$ (or $\bar{\Gamma}(N)$, if we are considering matrices modulo $\pm I$). Notice that a congruence subgroup of level N also has level N' for any multiple N' of N. This is because $\Gamma(N) \supset \Gamma(N')$. We say that a subgroup of Γ (or of $\bar{\Gamma}$) is a "congruence subgroup" if there exists N such that it is a congruence subgroup of level N. Not all subgroups of Γ are congruence

subgroups, but we shall never have occasion to deal with non-congruence subgroups.

For our purposes the most important congruence subgroups of Γ are:

$$\Gamma_0(N) \underset{\text{def}}{=} \left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma \middle| c \equiv 0 \ (\text{mod } N)\right\}; \tag{1.4}$$

$$\Gamma_1(N) \underset{\text{def}}{=} \left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) \middle| a \equiv 1 \ (\text{mod } N)\right\}. \tag{1.5}$$

It is easy to check that these are, in fact, subgroups. Note that if $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_1(N)$, then since $ad - bc = 1$ and $N|c$, it follows that $ad \equiv 1 \ (\text{mod } N)$, and hence $d \equiv 1 \ (\text{mod } N)$. We sometimes abbreviate the definitions (1.3)–(1.5) as follows:

$$\Gamma(N) = \left\{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{mod } N\right\}; \qquad \Gamma_1(N) = \left\{\begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \text{mod } N\right\};$$

$$\Gamma_0(N) = \left\{\begin{pmatrix} * & * \\ 0 & * \end{pmatrix} \text{mod } N\right\},$$

where $*$ indicates the absence of any congruence condition modulo N.

Whenever a group acts on a set, it divides the set into equivalence classes, where two points are said to be in the same equivalence class if there is an element of the group which takes one to the other. In particular, if G is a subgroup of Γ, we say that two points $z_1, z_2 \in H$ are "G-equivalent" if there exists $g \in G$ such that $z_2 = gz_1$.

Let F be a closed region in H. (Usually, F will also be simply connected.) We say that F is a "fundamental domain" for the subgroup G of Γ if every $z \in H$ is G-equivalent to a point in F, but no two distinct points z_1, z_2 in the *interior* of F are G-equivalent (two boundary points are permitted to be G-equivalent).

The most famous example of a fundamental domain is shown in Fig. III.1:

$$F \underset{\text{def}}{=} \{z \in H | -\tfrac{1}{2} \leq \text{Re } z \leq \tfrac{1}{2} \text{ and } |z| \geq 1\}. \tag{1.6}$$

Proposition 1. *The region F defined in* (1.6) *is a fundamental domain for* Γ.

PROOF. The group $\Gamma = SL_2(\mathbb{Z})$ contains two fractional linear transformations which act as building blocks for the entire group:

$$\begin{aligned} T &\underset{\text{def}}{=} \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}: z \mapsto z + 1; \\ S &\underset{\text{def}}{=} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}: z \mapsto -1/z. \end{aligned} \tag{1.7}$$

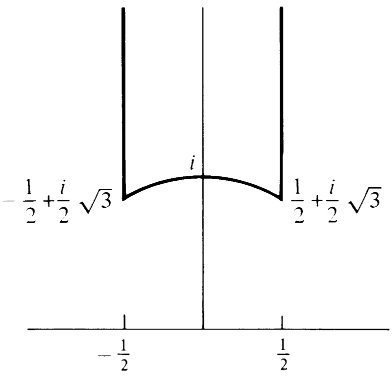


Figure III.1

To prove that every $z \in H$ is Γ-equivalent to a point in F, the idea is to use translations T^j to move a point z inside the strip $-\frac{1}{2} \leq \operatorname{Re} z \leq \frac{1}{2}$. If it lands outside the unit circle, it is in F. Otherwise use S to throw the point outside the unit circle, then use a translation T^k to bring it inside the strip, and continue in this way until you get a point inside the strip and outside the unit circle. We now give a precise proof.

Let $z \in H$ be fixed. Let Γ' be the subgroup of Γ generated by S and T (we shall soon see that actually $\Gamma = \pm\Gamma'$). If $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma'$, then $\operatorname{Im} \gamma z = \operatorname{Im} z/|cz + d|^2$ by (1.2). Since c and d are integers, the numbers $|cz + d|$ are bounded away from zero. (Geometrically, as c and d vary through all integers, the complex numbers $cz + d$ run through the lattice generated by 1 and z; and there is a disc around 0 which contains no nonzero lattice point.) Thus, there is some $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma'$ such that $\operatorname{Im} \gamma z$ is maximal. Replacing γ by $T^j\gamma$ for some suitable j, without loss of generality we may suppose that γz is in the strip $-\frac{1}{2} \leq \operatorname{Re} \gamma z \leq \frac{1}{2}$. But then, if γz were not in F, i.e., if we had $|\gamma z| < 1$, then, by (1.2), we would have:

$$\operatorname{Im} S\gamma z = \operatorname{Im} \gamma z/|\gamma z|^2 > \operatorname{Im} \gamma z,$$

which contradicts our choice of $\gamma \in \Gamma'$ so that $\operatorname{Im} \gamma z$ is maximal. Thus, there exists $\gamma \in \Gamma'$ such that $\gamma z \in F$.

We now prove that no two points in the interior of F are Γ-equivalent. We shall actually prove a more precise result. Suppose that $z_1, z_2 \in F$ are Γ-equivalent. Here we are not supposing that z_1 and z_2 are necessarily distinct or that they are necessarily in the interior of F. Without loss of generality, suppose that $\operatorname{Im} z_2 \geq \operatorname{Im} z_1$. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ be such that $z_2 = \gamma z_1$. Since $\operatorname{Im} z_2 \geq \operatorname{Im} z_1$, by (1.2) we must have $|cz_1 + d| \leq 1$. Since z_1 is in F and d is real (in fact, an integer), it is easy to see from Fig. III.1 that this inequality cannot hold if $|c| \geq 2$. This leaves the cases: (i) $c = 0$, $d = \pm 1$; (ii) $c = \pm 1$, $d = 0$, and z_1 is on the unit circle; (iii) $c = d = \pm 1$ and $z_1 = -\frac{1}{2} + \frac{\sqrt{-3}}{2}$; (iv) $c = -d = \pm 1$ and $z_1 = \frac{1}{2} + \frac{\sqrt{-3}}{2}$. In case (i), either γ or $-\gamma$

is a translation T^j; but such a γ can take a point in F to another point in F only if it is the identity or if $j = \pm 1$ and the points are on the two vertical boundary lines $\mathrm{Re}\, z = \pm\frac{1}{2}$. In case (ii), it is easy to see that $\gamma = \pm T^a S$ with $a = 0$ and z_1 and z_2 on the unit circle (and symmetrically located with respect to the imaginary axis) or with $a = \pm 1$ and $z_1 = z_2 = \pm\frac{1}{2} + \frac{\sqrt{-3}}{2}$. In case (iii), γ can be written as $\pm T^a\begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}$, and if such a map takes $z_1 \in F$ to $z_2 \in F$ we must have $a = 0$ and $z_2 = z_1 = -\frac{1}{2} + \frac{\sqrt{-3}}{2}$ or else $a = 1$ and $z_2 = z_1 + 1 = \frac{1}{2} + \frac{\sqrt{-3}}{2}$. Case (iv) is handled in the same way as case (iii). We conclude that in no case can z_1 or z_2 belong to the interior of F, unless $\pm\gamma$ is the identity and $z_2 = z_1$. This proves the proposition. □

In the course of the proof of Proposition 1, we have established two other facts.

Proposition 2. *Two distinct points z_1, z_2 on the boundary of F are Γ-equivalent only if* $\mathrm{Re}\, z_1 = \pm\frac{1}{2}$ *and* $z_2 = z_1 \pm 1$, *or if* z_1 *is on the unit circle and* $z_2 = \frac{-1}{z_1}$.

In the next proposition, we use the notation G_z for the "isotropy subgroup" of an element z in a set on which the group G acts: by definition, $G_z = \{g \in G \mid gz = z\}$.

Proposition 3. *If* $z \in F$, *then* $\Gamma_z = \pm I$ *except in the following three cases:*

(i) $\Gamma = \pm\{I, S\}$ *if* $z = i$;
(ii) $\Gamma = \pm\{I, ST, (ST)^2\}$ *if* $z = \omega = -\frac{1}{2} + \frac{\sqrt{-3}}{2}$;
(iii) $\Gamma = \pm\{I, TS, (TS)^2\}$ *if* $z = -\bar{\omega} = \frac{1}{2} + \frac{\sqrt{-3}}{2}$.

Both Propositions 2 and 3 follow from the second part of the proof of Proposition 1.

Notice that $S^2 = -I$, $(ST)^3 = -I$, and $(TS)^3 = -I$. Thus, in $\bar{\Gamma} = SL_2(\mathbb{Z})/\pm I$ the elements S, ST, TS generate cyclic subgroups of order 2, 3, 3, respectively; and these subgroups are the isotropy subgroups of the elements i, ω, $-\bar{\omega}$, respectively. These points in H with nontrivial isotropy subgroups are called "elliptic points".

Another by-product of the proof of Proposition 1 is the following useful fact.

Proposition 4. *The group* $\bar{\Gamma} = SL_2(\mathbb{Z})/\pm I$ *is generated by the two elements* S *and* T (*see* (1.7)). *In other words, any fractional linear transformation can be written as a "word" in* S (*the negative-reciprocal map*) *and* T (*translation by* 1) *and their inverses.*

PROOF. As in the proof of Proposition 1, let Γ' denote the subgroup of Γ generated by S and T. Let z be any point in the interior of F (e.g. $z = 2i$). Let g be an element of Γ. Consider the point $gz \in H$. In the first part of the

proof of Proposition 1, we showed that there exists $\gamma \in \Gamma'$ such that $\gamma(gz) \in F$. But since z is in the interior of F, it follows by Propositions 1 and 3 that $\gamma g = \pm I$, i.e., $g = \pm\gamma^{-1} \in \Gamma'$. This shows that any $g \in \Gamma$ is actually (up to a sign) in Γ'. The proposition is proved. □

Thus, any element in $\bar{\Gamma}$ can be written in the form $S^{a_1}T^{b_1}S^{a_2}T^{b_2} \cdots S^{a_l}T^{b_l}$, where all of the a_j, b_j are nonzero integers, except that we allow a_1 and/or b_l to be zero; and, since $S^2 = -I$, we may suppose that all of the a_j equal 1, except that $a_1 = 0$ or 1. We may also use the identity $(ST)^3 = -1$ to achieve a further simplification in some cases.

We now return to the fundamental domain F for $\Gamma = SL_2(\mathbb{Z})$. Recall that in Chapter I §4 we also had a fundamental domain, in that case a parallelogram $\Pi \subset \mathbb{C}$ for the lattice L. In that case the group was L, the action of $g \in L$ on a point $z \in \mathbb{C}$ was simply $g(z) = g + z$. Every $z \in \mathbb{C}$ is L-equivalent to a point in Π, and no two points in the interior of Π are L-equivalent. In that situation we found it useful to glue together the boundary of Π by identifying L-equivalent points. We obtained a torus, and then we found that the map $z \mapsto (\wp(z), \wp'(z))$ gives an analytic isomorphism from the torus $\mathbb{C}/L$ to the elliptic curve $y^2 = 4x^3 - g_2(L)x - g_3(L)$ (see §I.6).

In our present situation, with the group Γ acting on the set H with fundamental domain F, it is also useful to identify Γ-equivalent points on the boundary of F. Visually, we fold F around the imaginary axis, gluing the point $\frac{1}{2} + iy$ to $-\frac{1}{2} + iy$ and the point $e^{2\pi i\theta}$ to $e^{2\pi i(1/2-\theta)}$ for $y \geq \frac{\sqrt{3}}{2}$ and $\frac{1}{6} \leq \theta \leq \frac{1}{3}$. The resulting set F with its sides glued is in one-to-one correspondence with the set of Γ-equivalence classes in H, which we denote $\Gamma \backslash H$. (The notation $\Gamma\backslash H$ rather than H/Γ is customarily used because the group Γ acts on the set H on the left.)

In Chapter I we saw how useful it is to "complete" the picture by including a point or points at infinity. The same is true when we work with $\Gamma\backslash H$.

Let $\bar{H}$ denote the set $H \cup \{\infty\} \cup \mathbb{Q}$. That is, we add to H a point at infinity (which should be visualized far up the positive imaginary axis; for this reason we sometimes denote it $i\infty$) and also all of the *rational* numbers on the real axis. These points $\{\infty\} \cup \mathbb{Q}$ are called "cusps". It is easy to see that Γ permutes the cusps transitively. Namely, any fraction a/c in lowest terms can be completed to a matrix $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ by solving $ad - bc = 1$ for d and b; this matrix takes ∞ to a/c. Hence all rational numbers are in the same Γ-equivalence class as ∞.

If Γ' is a subgroup of Γ, then Γ' permutes the cusps, but in general not transitively. That is, there is usually more than one Γ'-equivalence class among the cusps $\{\infty\} \cup \mathbb{Q}$. We shall see examples later. By a "cusp of Γ'" we mean a Γ'-equivalence class of cusps. We may choose any convenient representative of the equivalence class to denote the cusp. Thus, we say that "Γ has a single cusp at ∞", where ∞ can be replaced by any rational number a/c in this statement.

We extend the usual topology on H to the set $\bar{H}$ as follows. First, a funda-

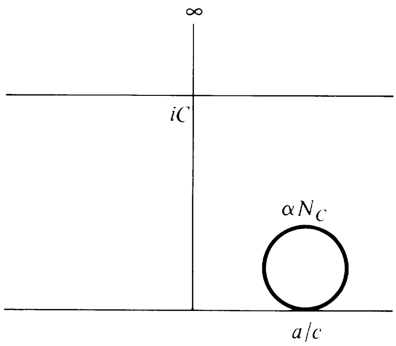


Figure III.2

mental system of open neighborhoods of ∞ is $N_C = \{z \in H | \operatorname{Im} z > C\} \cup \{\infty\}$ for any $C > 0$. Note that if we map H to the punctured open unit disc by sending

$$z \mapsto q \underset{\text{def}}{=} e^{2\pi i z}, \tag{1.8}$$

and if we agree to take the point $\infty \in \bar{H}$ to the origin under this map, then N_C is the inverse image of the open disc of radius $e^{-2\pi C}$ centered at the origin, and we have defined our topology on $H \cup \{\infty\}$ so as to make this map (1.8) continuous.

The change of variables (1.8) from z to q plays a basic role in the theory of modular functions. We use (1.8) to define an analytic structure on $H \cup \{\infty\}$. In other words, given a function on H of period 1, we say that it is meromorphic at ∞ if it can be expressed as a power series in the variable q having at most finitely many negative terms, i.e., it has a Fourier expansion of the form

$$f(z) = \sum_{n \in \mathbb{Z}} a_n e^{2\pi i n z} = \sum_{n \in \mathbb{Z}} a_n q^n, \tag{1.9}$$

in which $a_n = 0$ for $n \ll 0$. We say that $f(z)$ is holomorphic at ∞ if $a_n = 0$ for all negative n; and we say that $f(z)$ vanishes at ∞ if $f(z)$ is holomorphic at ∞ and $a_0 = 0$. More generally, if $f(z)$ has period N, then we use the map $z \mapsto q_N \underset{\text{def}}{=} e^{2\pi i z/N}$ to map $H \cup \{\infty\}$ to the open unit disc. We then express $f(z)$ as a series in q_N, and say that it is meromorphic (is holomorphic, vanishes) at ∞ if $a_n = 0$ for $n \ll 0$ (respectively, for $n < 0$, for $n \leq 0$).

Next, near a cusp $a/c \in \mathbb{Q} \subset \bar{H}$ we define a fundamental system of open neighborhoods by completing a, c to a matrix $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ and using α to transport the N_C to discs which are tangent to the real axis at a/c (see Fig. III.2). In other words, with this topology, to say that a sequence z_j approaches a/c means that $\alpha^{-1} z_j$ approaches $i\infty$, i.e., that $\operatorname{Im} \alpha^{-1} z_j$ approaches infinity in the usual sense. Notice, by the way, that the topology near the rational numbers a/c does *not* agree with the usual topology on the real line, i.e., a

sequence of rational numbers which approaches a/c as real numbers will *not* approach a/c in our topology.

Let $\bar{F}$ denote the fundamental domain F with Γ-equivalent boundary points identified and with the cusp ∞ thrown in. Thus, the points of $\bar{F}$ are in one-to-one correspondence with Γ-equivalence classes in $\bar{H}$. We take the topology on $\bar{F}$ which comes from the topology on $\bar{H}$. That is, by a small disc around an interior point of F we mean a disc in the usual sense; by a small disc around ∞ we mean all points lying about the line $\text{Im } z = C$, where C is large; by a small disc around a boundary point $-\frac{1}{2} + iy$ we mean the half-disc contained in F together with the half-disc of the same radius around $\frac{1}{2} + iy$ which is contained in F; and so on. Thus, $\bar{F}$ has an analytic structure coming from the usual structure on H, except at ∞, where it comes from the usual structure at 0 after the change of variable (1.8).

In Chapter I we found that the Weierstrass $\wp$-function gave us an analytic isomorphism of $\mathbb{C}/L$ with an elliptic curve in $\mathbb{P}^2_{\mathbb{C}}$. In our present situation, we shall later see that a certain function (called the "j-invariant") gives an analytic isomorphism between $\bar{F} = \Gamma \backslash \bar{H}$ and the projective line (Riemann sphere) $\mathbb{P}^1_{\mathbb{C}}$.

We now look at fundamental domains for subgroups $\Gamma' \subset \Gamma$. Suppose that $[\Gamma : \Gamma'] = n < \infty$, so that Γ can be written as a disjoint union of n cosets $\Gamma = \bigcup_{i=1}^{n} \alpha_i \Gamma'$. I claim that $F' = \bigcup_{i=1}^{n} \alpha_i^{-1} F$ will serve as a fundamental domain for Γ'. Let us verify that every $z \in H$ is Γ'-equivalent to a point in F'; the verification that no two interior points of F' are Γ'-equivalent will be left as an exercise (see below). Let $z \in H$. Since F is a fundamental domain for Γ, we can find $\gamma \in \Gamma$ such that $\gamma z \in F$. Then for some i we have $\gamma = \alpha_i \gamma'$ with $\gamma' \in \Gamma'$, and hence $\gamma' z \in \alpha_i^{-1} F \subset F'$, as desired. Roughly speaking, F' is n times as big as F because there are one n-th as many elements $\gamma' \in \Gamma'$ which one can use to move z around.

There are many possible choices of the α_i in the coset decomposition in the previous paragraph. In practice, we shall always try to choose the α_i so that the resulting F' is simply connected.

As an example, let us find a fundamental domain $F(2)$ for $\Gamma' = \Gamma(2)$. (We shall use the notation $F(N)$ to mean any fundamental domain for $\Gamma(N)$, $F_0(N)$ to mean any fundamental domain for $\Gamma_0(N)$, etc.) Since $\Gamma(2)$ is the kernel of the surjective map $SL_2(\mathbb{Z}) \to SL_2(\mathbb{Z}/2\mathbb{Z})$, and since $SL_2(\mathbb{Z}/2\mathbb{Z}) = GL_2(\mathbb{Z}/2\mathbb{Z})$ is isomorphic to S_3 (see Problem 6 of §I.8), it follows that $[\Gamma : \Gamma(2)] = 6$. As coset representatives of Γ modulo $\Gamma(2)$ let us choose:

$$\alpha_1 = I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}; \qquad \alpha_2 = T = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix};$$

$$\alpha_3 = S = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}; \qquad \alpha_4 = TS = \begin{pmatrix} 1 & -1 \\ 1 & 0 \end{pmatrix};$$

$$\alpha_5 = ST = \begin{pmatrix} 0 & -1 \\ 1 & 1 \end{pmatrix}; \qquad \alpha_6 = T^{-1}ST = \begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}.$$

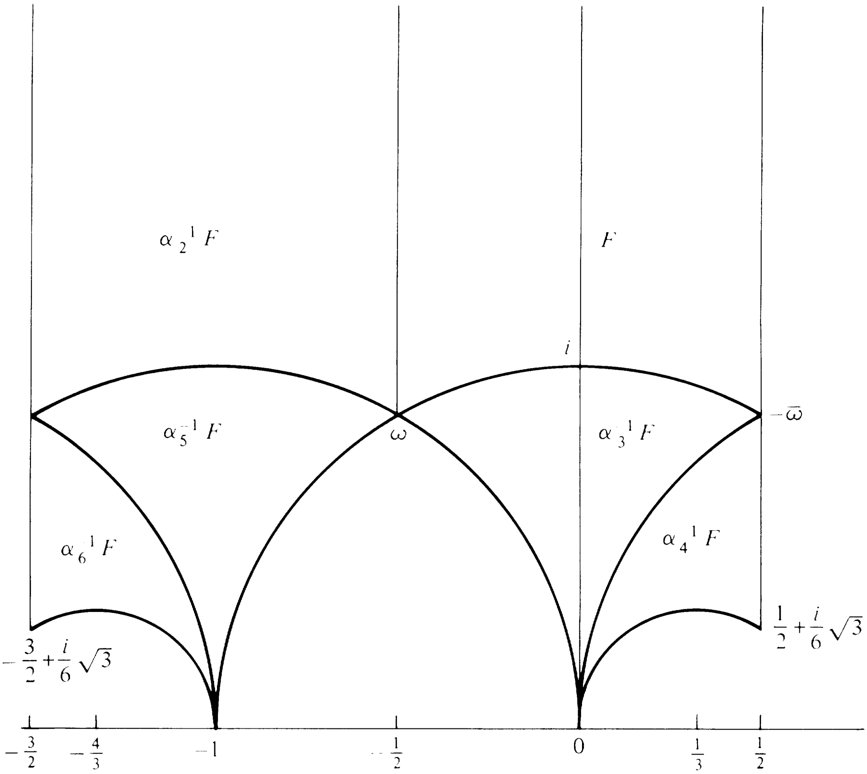


Figure III.3. A fundamental domain $F(2)$ for $\Gamma(2)$.

The resulting fundamental domain $F(2) = \bigcup_{i=1}^{6} \alpha_i^{-1}F$ is depicted in Fig. III.3. Because any fractional linear transformation takes a circle or line to a circle or line and preserves symmetry about the real axis, it follows that the boundary of any fundamental domain constructed in this way consists of vertical lines and arcs of circles centered at rational numbers on the real axis. The boundary of the fundamental domain in Fig. III.3 is made up from the vertical lines $\operatorname{Re} z = -\frac{3}{2}$ and $\operatorname{Re} z = \frac{1}{2}$, the circles of radius 1 centered at 0 and -1, and the circles of radius $\frac{1}{3}$ centered at $\frac{1}{3}$ and $-\frac{4}{3}$.

We see that $\Gamma(2)$ has three cusps: ∞, 0, -1. That is, there are three $\Gamma(2)$-equivalence classes of cusps with those three points as representatives. Now the etymology of the word "cusp" is clear: geometrically, $F(2)$ has the appearance we usually associate with the word "cusp" at the points 0 and -1.

Problems

1. Prove that $\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$ but not of Γ. Is $\Gamma_0(N)$ a normal subgroup of Γ?

2. (a) Prove that for any N, the map $SL_2(\mathbb{Z}) \to SL_2(\mathbb{Z}/N\mathbb{Z})$ obtained by reducing the matrix entries modulo N is a *surjective* group homomorphism.
   (b) Prove that for any positive integers M and N, the maps ("reduction modulo N") from $SL_2(\mathbb{Z}/MN\mathbb{Z})$ to $SL_2(\mathbb{Z}/N\mathbb{Z})$ and from $GL_2(\mathbb{Z}/MN\mathbb{Z})$ to $GL_2(\mathbb{Z}/N\mathbb{Z})$ are surjective group homomorphisms.

3. What is the order of the group (a) $GL_2(\mathbb{F}_q)$? (b) $SL_2(\mathbb{F}_q)$?

4. (a) What is the kernel of the homomorphism $GL_2(\mathbb{Z}/p^e\mathbb{Z}) \to GL_2(\mathbb{Z}/p\mathbb{Z})$?
   (b) What is the order of the group $GL_2(\mathbb{Z}/p^e\mathbb{Z})$?
   (c) What is the order of the group $SL_2(\mathbb{Z}/p^e\mathbb{Z})$?

5. Let $N = p_1^{e_1} \cdots p_r^{e_r}$ be the prime factorization of the positive integer N. Show that the reductions modulo $p_j^{e_j}$, $j = 1, \ldots, r$, give isomorphisms

$$GL_2(\mathbb{Z}/N\mathbb{Z}) \xrightarrow{\sim} \Pi_j GL_2(\mathbb{Z}/p_j^{e_j}\mathbb{Z}) \quad \text{and} \quad SL_2(\mathbb{Z}/N\mathbb{Z}) \xrightarrow{\sim} \Pi_j SL_2(\mathbb{Z}/p_j^{e_j}\mathbb{Z}).$$

6. What is the order of the group $SL_2(\mathbb{Z}/N\mathbb{Z})$?

7. Find the indices $[\Gamma\colon \Gamma(N)]$, $[\Gamma_1(N)\colon \Gamma(N)]$, $[\Gamma_0(N)\colon \Gamma_1(N)]$, $[\Gamma_0(N)\colon \Gamma(N)]$, and $[\Gamma\colon \Gamma_0(N)]$.

8. Find a simple isomorphism from $\Gamma(N)$ to a subgroup of $\Gamma_0(N^2)$ having index $\phi(N)$ in $\Gamma_0(N^2)$. In particular, $\Gamma(2)$ and $\Gamma_0(4)$ are isomorphic.

9. Suppose that Γ_1 and Γ_2 are two subgroups of finite index in Γ, and $\Gamma_1 = \alpha\Gamma_2\alpha^{-1}$ for some $\alpha \in GL_2(\mathbb{Q})$. If F_2 is a fundamental domain for Γ_2, prove that αF_2 is a fundamental domain for Γ_1.

10. Using the previous problem, draw a fundamental domain for $\Gamma_0(4)$.

11. Suppose that $\bar{\Gamma} = \bigcup_{j=1}^n \alpha_j \bar{\Gamma}'$, where $\bar{\Gamma}'$ is a subgroup of index n in $\bar{\Gamma}$. Let $F' = \bigcup_{j=1}^n \alpha_j^{-1} F$. Show that no two distinct points in the interior of F' are $\bar{\Gamma}'$-equivalent.

12. Describe all congruence subgroups Γ' of level 2, i.e., all groups contained between Γ and $\Gamma(2)$: $\Gamma(2) \subset \Gamma' \subset \Gamma$. For each such Γ', find a simply connected fundamental domain by taking a suitable part of the fundamental domain $F(2)$ for $\Gamma(2)$ that was given in the text.

13. (a) Prove that $\pm S$ and T^2, with S and T as in (1.7), generate one of the subgroups of level 2 in the previous problem. That group was denoted $\mathfrak{G}(2)$ by Hecke.
   (b) Prove that TST and T^2 generate $\bar{\Gamma}^0(2) \underset{\text{def}}{=} \{(\begin{smallmatrix} * & 0 \\ * & * \end{smallmatrix}) \bmod 2\}$; and that T and ST^2S generate $\bar{\Gamma}_0(2)$.
   (c) Prove that the elements T^2 and $ST^{-2}S$ generate $\bar{\Gamma}(2)$.
   (d) Prove that T and $ST^{-4}S$ generate $\bar{\Gamma}_0(4)$.

14. (a) Prove that the following is a complete set of coset representatives $\{\alpha_j\}$ for $\Gamma_0(p^e)$ in Γ, i.e., $\Gamma = \bigcup \alpha_j \Gamma_0(p^e)$ is a disjoint union:

$$1; \qquad T^{-k}S, \quad k = 0, 1, \ldots, p^e - 1; \qquad ST^{-kp}S, \quad k = 1, 2, \ldots, p^{e-1} - 1.$$

   (b) Use part (a) to draw a fundamental domain for $\Gamma_0(4)$.
   (c) Using part (a), describe a fundamental domain for $\Gamma_0(p)$. Draw the fundamental domain for $\Gamma_0(3)$.

15. Suppose Γ_1 and Γ_2 are subgroups of finite index in Γ having fundamental domains

F_1 and F_2, respectively. Does $F_1 \subset F_2$ imply that $\Gamma_1 \supset \Gamma_2$? (Give a proof or counterexample.)

16. (a) The assertion at the end of the section that $\Gamma(2)$ has 3 cusps (i.e., that there are three $\Gamma(2)$-equivalence classes of cusps) does not immediately follow from the appearance of the fundamental domain, because boundary points of the fundamental domain *may* be $\Gamma(2)$-equivalent. But prove directly that ∞, 0, -1 are $\Gamma(2)$-inequivalent to one another.
(b) How many cusps does each of the congruence subgroups in Problem 12 have?

17. Let $\{\alpha_j\}$ be a complete set of coset representatives for $\bar{\Gamma}'$ in $\bar{\Gamma}$, where Γ' is a subgroup of finite index in Γ. Show that the cusps of Γ' are among the set $\{\alpha_j^{-1}\infty\}$, but that $\alpha_j^{-1}\infty$ and $\alpha_k^{-1}\infty$ are Γ'-equivalent if and only if there exists $n \in \mathbb{Z}$ such that

$$\alpha_k^{-1} T^n \alpha_j \in \bar{\Gamma}'.$$

18. Prove that $\Gamma_0(p)$ has two cusps ∞ and 0; and that $\Gamma_0(p^2)$ has $p + 1$ cusps: ∞, 0, and $-1/kp$ for $k = 1, \ldots, p - 1$.

§2. Modular forms for $SL_2(\mathbb{Z})$

Definition. Let $f(z)$ be a meromorphic function on the upper half-plane H, and let k be an integer. Suppose that $f(z)$ satisfies the relation

$$f(\gamma z) = (cz + d)^k f(z) \quad \text{for all} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL_2(\mathbb{Z}). \tag{2.1}$$

In particular, for the elements $\gamma = T = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$ and $\gamma = S = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$, (2.1) gives

$$f(z + 1) = f(z); \tag{2.2}$$

$$f(-1/z) = (-z)^k f(z). \tag{2.3}$$

Further suppose that $f(z)$ is "meromorphic at infinity". Recall that this means that the Fourier series

$$f(z) = \sum_{n \in \mathbb{Z}} a_n q^n, \quad \text{where} \quad q = e^{2\pi i z}, \tag{2.4}$$

has at most finitely many nonzero a_n with $n < 0$. Then $f(z)$ is called a "modular function of weight k for $\Gamma = SL_2(\mathbb{Z})$".

If, in addition, $f(z)$ is actually *holomorphic* on H and at infinity (i.e., $a_n = 0$ for all $n < 0$), then $f(z)$ is called a "modular form of weight k for $\Gamma = SL_2(\mathbb{Z})$". The set of such functions is denoted $M_k(\Gamma)$.

If we further have $a_0 = 0$, i.e., the modular form vanishes at infinity, then $f(z)$ is called a "cusp-form of weight k for Γ". The set of such functions is denoted $S_k(\Gamma)$. (The use of the letter S is traditional, and comes from the German "Spitzenform" for "cusp-form". Cusp-forms are sometimes also called "parabolic forms".)

Finally, the expansion (2.4) for a modular function $f(z)$ is called its "q-expansion".

We first make some easily verified remarks about these definitions.

Remarks. 1. If k is odd, there are no nonzero modular functions of weight k for Γ. We see this by substituting $\gamma = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}$ in (2.1). So in this section we shall always suppose that k is even.

2. Since $\frac{d\gamma z}{dz} = \frac{d}{dz}((az+b)/(cz+d)) = (cz+d)^{-2}$, we can rewrite (2.1) in the form: $(\frac{d\gamma z}{dz})^{k/2} f(\gamma z) = f(z)$, i.e., $f(z)(dz)^{k/2}$ is invariant when z is replaced by γz. From this we see that if (2.1) holds for γ_1 and γ_2, then it holds for $\gamma_1\gamma_2$. Since all of $\bar{\Gamma}$ is generated by S and T, this means that (2.2)–(2.3) imply (2.1).

3. The conditions are preserved under addition and scalar multiplication, i.e., the sets of modular functions, forms, and cusp-forms of some fixed weight are complex vector spaces. In addition, the product of a modular function (or form) of weight k_1 and a modular function (respectively, form) of weight k_2 is a modular function (respectively, form) of weight $k_1 + k_2$; and the quotient of a modular function of weight k_1 by a nonzero modular function of weight k_2 is a modular function of weight $k_1 - k_2$. In particular, the set of modular functions of weight zero is a field.

4. If $k = 0$, then the condition (2.1) says that $f(z)$ is invariant under $\Gamma = SL_2(\mathbb{Z})$, i.e., it may be considered as a function on $\Gamma\backslash H$. If $k = 2$, then the differential form $f(z)dz$ on H is invariant under Γ, i.e., $f(\gamma z)d\gamma z = f(z)dz$, since $d\gamma z/dz = (cz+d)^{-2}$.

Example (Eisenstein series). Let k be an even integer greater than 2. For $z \in H$ we define

$$G_k(z) \underset{\text{def}}{=} \sum_{m,n}{}' \frac{1}{(mz+n)^k}, \tag{2.5}$$

where the summation is over pairs of integers m, n not both zero. If we let L_z denote the lattice in $\mathbb{C}$ spanned by 1 and z, then this is a familiar definition from Chapter I (see (6.3) of Chapter I). That is, the function $G_k(z)$ in (2.5) is what we then denoted $G_k(L_z)$. The new point of view in this chapter is to think of $G_k(z) = G_k(L_z)$ as functions of z, not merely as coefficients in the Laurent expansion of the Weierstrass $\wp$-function. (The letter z was used in a different way in §6 of Chapter I.)

Because k is at least 4, the double sum (2.5) is absolutely convergent, and uniformly convergent in any compact subset of H. Hence $G_k(z)$ is a holomorphic function on H. It is also obvious that $G_k(z) = G_k(z+1)$, and that the Fourier expansion (2.4) has no negative terms, because $G_k(z)$ approaches a finite limit as $z \to i\infty$:

$$\lim_{z \to i\infty} \sum_{m,n}{}' (mz+n)^{-k} = \sum_{n \neq 0} n^{-k} = 2\zeta(k).$$

Finally, we easily check that

$$z^{-k}G_k(-1/z) = \sum_{m,n}{}' (-m + nz)^{-k} = G_k(z),$$

i.e., $G_k(z)$ satisfies (2.3). We have proved

Proposition 5.

$$G_k \in M_k(\Gamma).$$

We now compute the q-expansion coefficients for G_k. We shall find that these coefficients are essentially the arithmetic functions of n

$$\sigma_{k-1}(n) \underset{\text{def}}{=} \sum_{d|n} d^{k-1}. \tag{2.6}$$

Proposition 6. *Let k be an even integer greater than 2, and let $z \in H$. Then the modular form $G_k(z)$ defined in (2.5) has q-expansion*

$$G_k(z) = 2\zeta(k)\left(1 - \frac{2k}{B_k}\sum_{n=1}^{\infty} \sigma_{k-1}(n)q^n\right), \tag{2.7}$$

where $q = e^{2\pi i z}$, and the Bernoulli numbers B_k are defined by setting

$$\frac{x}{e^x - 1} = \sum_{k=0}^{\infty} B_k \frac{x^k}{k!}. \tag{2.8}$$

(*Note*. Our notation is slightly different from Serre's in *A Course in Arithmetic*. Basically, he uses $2k$ where we use k.)

Proof. The logarithmic derivative of the product formula for sine is

$$\pi \cot(\pi a) = \frac{1}{a} + \sum_{n=1}^{\infty}\left(\frac{1}{a+n} + \frac{1}{a-n}\right), \qquad a \in H. \tag{2.9}$$

If we write the left side as $\pi i(e^{\pi i a} + e^{-\pi i a})/(e^{\pi i a} - e^{-\pi i a}) = \pi i + 2\pi i/(e^{2\pi i a} - 1)$, multiply both sides by a, replace $2\pi i a$ by x, and expand both sides in powers of x (using (2.8)), we obtain the well-known formula for $\zeta(k)$:

$$\zeta(k) = -\frac{(2\pi i)^k}{2}\frac{B_k}{k!} \quad \text{for} \quad k > 0 \text{ even}. \tag{2.10}$$

Next, if we successively differentiate both sides of (2.9) with respect to a (see Problem 8(d)–(e) of §II.4) and then replace a by mz, we obtain:

$$\sum_{n=-\infty}^{\infty} \frac{1}{(mz+n)^k} = \frac{(2\pi i)^k}{(k-1)!}\sum_{n=1}^{\infty} n^{k-1}e^{2\pi i n m z} = -\frac{2k}{B_k}\zeta(k)\sum_{d=1}^{\infty} d^{k-1}q^{dm}$$

(where we have used (2.10) and replaced n by d and $e^{2\pi i z}$ by q). Thus,

$$\begin{aligned} G_k(z) &= 2\zeta(k) + 2\sum_{m=1}^{\infty}\sum_{n=-\infty}^{\infty}\frac{1}{(mz+n)^k} \\ &= 2\zeta(k)\left(1 - \frac{2k}{B_k}\sum_{m,d=1}^{\infty} d^{k-1}q^{dm}\right). \end{aligned}$$

Collecting coefficients of a fixed power q^n in the last double sum, we obtain the sum in (2.6) as the coefficient of q^n. This completes the proof. □

Because of Proposition 6, it is useful to define the "normalized Eisenstein series", obtained by dividing $G_k(z)$ by the constant $2\zeta(k)$ in (2.7):

$$E_k(z) = \frac{1}{2\zeta(k)} G_k(z) = 1 - \frac{2k}{B_k} \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n. \tag{2.11}$$

Thus, $E_k(z)$ is defined so as to have *rational* q-expansion coefficients. The first few E_k are:

$$E_4(z) = 1 + 240 \sum_{n=1}^{\infty} \sigma_3(n) q^n;$$

$$E_6(z) = 1 - 504 \sum_{n=1}^{\infty} \sigma_5(n) q^n;$$

$$E_8(z) = 1 + 480 \sum_{n=1}^{\infty} \sigma_7(n) q^n;$$

$$E_{10}(z) = 1 - 264 \sum_{n=1}^{\infty} \sigma_9(n) q^n;$$

$$E_{12}(z) = 1 + \frac{65520}{691} \sum_{n=1}^{\infty} \sigma_{11}(n) q^n;$$

$$E_{14}(z) = 1 - 24 \sum_{n=1}^{\infty} \sigma_{13}(n) q^n.$$

An alternate way of defining the normalized Eisenstein series is to sum only over relatively prime pairs m, n in (2.5):

$$E_k(z) = \frac{1}{2} \sum_{\substack{m,n \in \mathbb{Z} \\ (m,n)=1}} \frac{1}{(mz+n)^k}, \tag{2.12}$$

where (m, n) denotes the greatest common divisor. The equivalence of (2.12) and (2.11) will be left as an exercise (see below).

Example (The discriminant modular form $\Delta(z)$). Recall from our study of the Weierstrass $\wp$-function of a lattice L that one defines $g_2(L) = 60G_4(L)$, $g_3(L) = 140G_6(L)$ to be the coefficients occurring in the differential equation satisfied by the $\wp$-function (i.e., in the equation of the elliptic curve corresponding to L; see (6.8)–(6.9) in §I.6). Now let us define $g_2(z) = g_2(L_z)$, $g_3(z) = g_3(L_z)$ for $z \in H$. That is,

$$g_2(z) = 60G_4(z); \qquad g_3(z) = 140G_6(z). \tag{2.13}$$

Then $g_2(z)$ and $g_3(z)$ are modular forms for Γ of weight 4 and 6, respectively. Since $\zeta(4) = \pi^4/90$ and $\zeta(6) = \pi^6/945$, we can express g_2 and g_3 in terms of the normalized Eisenstein series E_4 and E_6 as follows:

$$g_2(z) = \tfrac{4}{3}\pi^4 E_4(z); \qquad g_3(z) = \tfrac{8}{27}\pi^6 E_6(z). \tag{2.14}$$

The discriminant of the elliptic curve corresponding to L_z—a function of the lattice which is nonvanishing for all nontrivial lattices, i.e., in all cases when the cubic polynomial $4x^3 - g_2(L_z)x - g_3(L_z)$ has distinct roots—is given by (see Problem 2 of §I.6)

$$\Delta(z) = g_2(z)^3 - 27g_3(z)^2 = \frac{(2\pi)^{12}}{1728}(E_4(z)^3 - E_6(z)^2). \tag{2.15}$$

By Remark 3 following the definition of a modular form, we see that $\Delta(z)$ is a modular form of weight 12 for Γ. Moreover, because both $E_4(z)$ and $E_6(z)$ have constant term $a_0 = 1$ in their q-expansions, we see that the constant term in (2.15) is zero, i.e., $\Delta(z)$ is a cusp-form. It is the first example of a cusp-form that we have seen. We shall later see that it is the cusp-form of lowest possible weight for Γ.

The q-expansion of $(2\pi)^{-12}\Delta(z)$ has rational coefficients, and the first coefficient is easily computed to be 1 (use: $E_4 = 1 + 240q + \cdots$, $E_6 = 1 - 504q + \cdots$). We shall later prove a remarkable product formula, due to Jacobi, for this discriminant function.

The example of Eisenstein series gives us one modular form for every even weight starting with 4. It might seem unfortunate to have to pass up 2. Is there any Eisenstein series that can be salvaged in the case $k = 2$?

It turns out that there is, but the normalized Eisenstein series E_2 is not a modular form. We use the same definition as for the other E_k, except that the double sum when $k = 2$ is not absolutely convergent, and we must take care with the order of summation. Note that in the proof of Proposition 5, when showing that $z^{-k}G_k(-1/z) = G_k(z)$, we needed to reverse the order of summation over m and n. Because this change of order of summation is no longer justified when we only have conditional convergence, it turns out that E_2 fails to transform "correctly" under $z \mapsto -1/z$. E_2 satisfies a slightly more complicated transformation rule, a rule that will be used in important ways later.

Thus, we define (here $'$ means that $n \neq 0$ if $m = 0$)

$$\begin{aligned} E_2(z) &= \frac{1}{2\zeta(2)} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty}{}' \frac{1}{(mz+n)^2} \\ &= 1 + \frac{3}{\pi^2} \sum_{m\neq 0} \sum_{n=-\infty}^{\infty} \frac{1}{(mz+n)^2} \\ &= 1 + \frac{6}{\pi^2} \sum_{m=1}^{\infty} \sum_{n=-\infty}^{\infty} \frac{1}{(mz+d)^2}. \end{aligned} \tag{2.16}$$

The inner sums clearly converge for any $z \in H$ and any m; and, once we obtain a different expression for the inner sums, we shall see that the outer sum over m then converges absolutely. As in the proof of Proposition 6,

we find that for any fixed $m = 1, 2, 3, \ldots$

$$\sum_{n=-\infty}^{\infty} \frac{1}{(mz+n)^2} = -\frac{4}{B_2}\zeta(2)\sum_{d=1}^{\infty} dq^{dm}, \quad \text{where} \quad q = e^{2\pi i z}.$$

This gives us

$$E_2 = 1 - 24 \sum_{m=1}^{\infty}\sum_{d=1}^{\infty} dq^{dm}.$$

Since $|q| < 1$, it is easy to see that the double sum over m and d is absolutely convergent. Collecting coefficients of q^n by summing over all divisors d of n, we obtain:

$$E_2 = 1 - 24 \sum_{n=1}^{\infty} \sigma_1(n) q^n. \tag{2.17}$$

As in the case of the higher E_k, E_2 is a holomorphic function on H which is periodic of period 1 and holomorphic at infinity. In order for it to be a modular form of weight 2 for Γ, all we would need in addition is for $z^{-2}E_2(-1/z)$ to equal $E_2(z)$. From (2.16) we find that

$$\begin{aligned} z^{-2}E_2(-1/z) &= \frac{1}{2\zeta(2)} \sum_{m=-\infty}^{\infty} \sideset{}{'}\sum_{n=-\infty}^{\infty} \frac{1}{(-m+nz)^2} \\ &= 1 + \frac{3}{\pi^2} \sum_{n=-\infty}^{\infty} \sum_{m \neq 0} \frac{1}{(mz+n)^2}. \end{aligned} \tag{2.18}$$

Thus, the extent to which E_2 fails to satisfy the "right" rule is a reflection of the alteration produced by reversing the order of summation. We now compute this "error term".

Proposition 7.

$$z^{-2}E_2(-1/z) = E_2(z) + \frac{12}{2\pi i z} \tag{2.19}$$

PROOF. The proposition says that $12/2\pi i z$ is the difference between the double sum (2.18) and the double sum (2.16). Suppose we introduce a "correction term" $a_{m,n}$ inside both double sums which causes the double sums to be absolutely convergent. In that case, the order of summation of the "corrected" double sums does not make any difference, i.e., the two "corrected" double sums are equal. It then follows that the difference between (2.18) and (2.16) is equal to the difference between $\Sigma_m \Sigma_n a_{m,n}$ and $\Sigma_n \Sigma_m a_{m,n}$. The idea is to choose $a_{m,n}$ to be a term close to $(mz+n)^{-2}$ but which is easier to sum.

Let us take

$$a_{m,n} = a_{m,n}(z) = \frac{1}{(mz+n-1)(mz+n)} = \frac{1}{(mz+n-1)} - \frac{1}{(mz+n)}.$$

Since the difference between $(mz+n)^{-2}$ and $a_{m,n}$ is $1/(mz+n)^2(mz+n-1)$, it follows that the following "corrected" E_2 is absolutely convergent:

$$\tilde{E}_2(z) = 1 + \frac{3}{\pi^2}\sum_{m\neq 0}\sum_{n=-\infty}^{\infty}\left(\frac{1}{(mz+n)^2} - a_{m,n}(z)\right)$$
$$= 1 + \frac{3}{\pi^2}\sum_{m\neq 0}\sum_{n=-\infty}^{\infty}\frac{1}{(mz+n)^2} + \frac{3}{\pi^2}\sum_{m\neq 0}\sum_{n=-\infty}^{\infty}\left(\frac{1}{mz+n} - \frac{1}{mz+n-1}\right).$$

But since the last inner sum telescopes to zero, we have: $\tilde{E}_2(z) = E_2(z)$. Because the double sum for $\tilde{E}_2$ is absolutely convergent, we have

$$\begin{aligned} E_2(z) &= 1 + \frac{3}{\pi^2}\sum_{n=-\infty}^{\infty}\sum_{m\neq 0}\left(\frac{1}{(mz+n)^2} - a_{m,n}(z)\right) \\ &= z^{-2}E_2(-1/z) - \frac{3}{\pi^2}\sum_{n=-\infty}^{\infty}\sum_{m\neq 0} a_{m,n}(z). \end{aligned} \tag{2.20}$$

So it remains to evaluate this last double sum.

Now this double sum differs from the sum in (2.18) by an absolutely convergent series. As in the derivation of (2.17), we find that for $n > 0$

$$\sum_{m\neq 0}\frac{1}{(mz-n)^2} = \frac{1}{z^2}\sum_{m\neq 0}\frac{1}{(-n/z+m)^2} = -\frac{1}{n^2} - \frac{4\pi^2}{z^2}\sum_{d=1}^{\infty} de^{-2\pi i d n/z}.$$

Since $-1/z$ is a fixed element of H, we see that the outer sum over n in (2.18) converges absolutely. Thus, the same is true for $\Sigma_n\Sigma_m a_{m,n}(z)$. This justifies writing

$$\begin{aligned} \sum_{n=-\infty}^{\infty}\sum_{m\neq 0} a_{m,n}(z) &= \lim_{N\to\infty}\sum_{n=-N+1}^{N}\sum_{m\neq 0} a_{m,n}(z) \\ &= \lim_{N\to\infty}\sum_{m\neq 0}\sum_{n=-N+1}^{N} a_{m,n}(z). \end{aligned}$$

The last inner sum telescopes to $1/(mz-N) - 1/(mz+N)$, and by (2.9) we have

$$\begin{aligned} \sum_{m\neq 0}\left(\frac{1}{mz-N} - \frac{1}{mz+N}\right) &= \frac{2}{z}\sum_{m=1}^{\infty}\left(\frac{1}{-N/z+m} + \frac{1}{-N/z-m}\right) \\ &= \frac{2}{z}\left(\pi\cot\left(-\frac{\pi N}{z}\right) + \frac{z}{N}\right). \end{aligned}$$

We conclude that the double sum is equal to

$$\frac{2\pi}{z}\lim_{N\to\infty}\cot(-\pi N/z) = \frac{2\pi}{z}\lim_{N\to\infty} i\frac{e^{-2\pi i N/z}+1}{e^{-2\pi i N/z}-1} = -\frac{2\pi i}{z}.$$

Substituting this into (2.20) gives:

$$E_2(z) = z^{-2}E_2(-1/z) + \frac{6i}{\pi z} = z^{-2}E_2(-1/z) - \frac{12}{2\pi i z}.$$

This completes the proof of Proposition 7. □

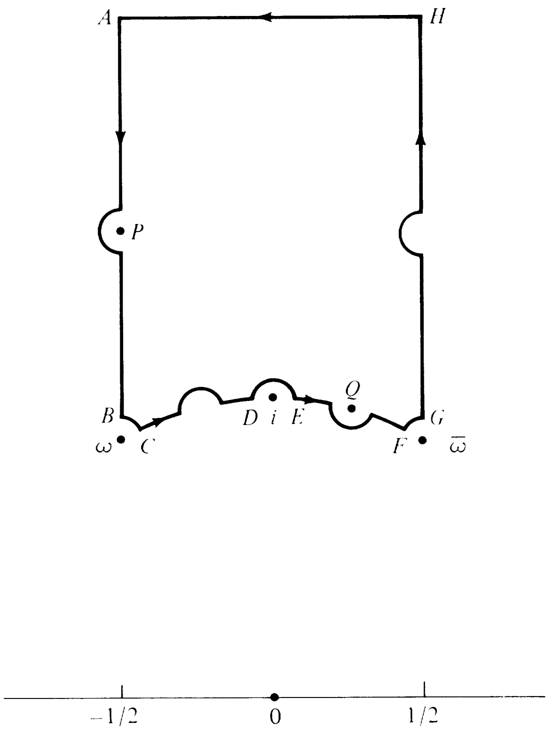


Figure III.4. Contour for the proof of Proposition 8.

The next result will play a basic role in determining the spaces $M_k(\Gamma)$ and $S_k(\Gamma)$ of modular forms and cusp-forms of given weight for Γ, and it will also be useful in proving that two modular forms defined in different ways are actually the same in certain cases.

Proposition 8. *Let $f(z)$ be a nonzero modular function of weight k for Γ. For $P \in H$, let $v_P(f)$ denote the order of zero (or minus the order of pole) of $f(z)$ at the point P. Let $v_\infty(f)$ denote the index of the first nonvanishing term in the q-expansion of $f(z)$. Then*

$$v_\infty(f) + \frac{1}{2}v_i(f) + \frac{1}{3}v_\omega(f) + \sum_{P \in \Gamma\backslash H, P \neq i, \omega} v_P(f) = \frac{k}{12}. \tag{2.21}$$

(*Note.* It is easy to check that $v_P(f)$ does not change if P is replaced by γP for $\gamma \in \Gamma$.)

PROOF. The idea of the proof is to count the zeros and poles in $\Gamma\backslash H$ by integrating the logarithmic derivative of $f(z)$ around the boundary of the fundamental domain F. More precisely, let C be the contour in Fig. III.4. The top of C is a horizontal line from $H = \frac{1}{2} + iT$ to $A = -\frac{1}{2} + iT$, where T is taken larger than the imaginary part of any of the zeros or poles of $f(z)$.

(*Note*. That this can be done, i.e., that $f(z)$ does not have poles or zeros with arbitrarily large imaginary part, follows from the fact that the change of variables $q = e^{2\pi iz}$ makes $f(z)$ into a meromorphic function of q in a disc around $q = 0$.) The rest of the contour follows around the boundary of F, except that it detours around any zero or pole on the boundary along circular arcs of small radius ε. This is done in such a way as to include every Γ-equivalence class of zero or pole exactly once inside C, except that i and ω (and $S\omega = -\bar{\omega}$) are kept outside of C if they are zeros or poles. In Fig. III.4 we have illustrated the case when the zeros and poles on the boundary of F consist of i, ω (and hence also $S\omega$), one point P on the vertical boundary line (and hence also the Γ-equivalent point on the opposite line), and one point Q on the unit circle part of the boundary (and hence also SQ).

According to the residue theorem, we have

$$\frac{1}{2\pi i}\int_C \frac{f'(z)}{f(z)}dz = \sum_{P\in\Gamma\backslash H,\, P\neq i,\omega} v_P(f). \tag{2.22}$$

On the other hand, we evaluate the integral in (2.22) section by section. First of all, the integral from A to B (see Fig. III.4) cancels the integral from G to H, because $f(z+1) = f(z)$, and the lines go in opposite directions. Next, we evaluate the integral over HA. To do this we make the change of variables $q = e^{2\pi iz}$. Let $\tilde{f}(q) = f(z) = \Sigma a_n q^n$ be the q-expansion. Since $f'(z) = \frac{d}{dq}\tilde{f}(q)\frac{dq}{dz}$, we find that this section of the integral in (2.22) is equal to the following integral over the circle of radius $e^{-2\pi T}$ centered at zero:

$$\frac{1}{2\pi i}\int \frac{d\tilde{f}/dq}{\tilde{f}(q)}dq.$$

Since the circle is traversed in a clockwise direction as z goes from H to A, it follows that this integral is minus the order of zero or pole of $\tilde{f}(q)$ at 0, and this is what we mean by $-v_\infty(f)$.

To evaluate the integral over the arcs BC, DE, and FG, recall the derivation of the residue formula. If $f(z)$ has Laurent expansion $c_m(z-a)^m + \cdots$ near a, with $c_m \neq 0$, then $f'(z)/f(z) = \frac{m}{z-a} + g(z)$, with $g(z)$ holomorphic at a. If we integrate $f'(z)/f(z)$ counterclockwise around a circular arc of angle θ centered at a with small radius ε, then as $\varepsilon \to 0$ this integral approaches $mi\theta$ (the usual residue formula results when $\theta = 2\pi$). We apply this to the section of (2.22) between B and C, letting $\varepsilon \to 0$. The angle approaches $\pi/3$, and so we obtain $-\frac{1}{2\pi i}(v_\omega(f)i\pi/3) = -v_\omega(f)/6$. (The minus sign is because the arc BC goes clockwise.) In the same way, we find that as $\varepsilon \to 0$ the part of (2.22) from D to E becomes $-v_i(f)/2$, and the part between F and G becomes $-v_{-\bar{\omega}}(f)/6 = -v_\omega(f)/6$.

What remains is the integral from C to D and from E to F. Combining the above calculations, we find from (2.22) that the left side of (2.21) is equal to that remaining section of the left side of (2.22). Thus, to prove Proposition 8, it remains to show that in the limit as $\varepsilon \to 0$

$$\frac{1}{2\pi i}\int_{CD}\frac{f'(z)}{f(z)}dz+\frac{1}{2\pi i}\int_{EF}\frac{f'(z)}{f(z)}dz\to\frac{k}{12}. \tag{2.23}$$

To compute the sum of these two integrals, we note that the transformation $S\colon z\mapsto -1/z$ takes CD to EF, or more precisely, to FE, i.e., Sz goes from F to E along the contour as z goes from C to D along the contour. The desired formula (2.23) will follow from the following more general lemma.

Lemma. *Let $\gamma=\left(\begin{smallmatrix}a&b\\c&d\end{smallmatrix}\right)\in\Gamma$, with $c\neq 0$, and let $f(z)$ be a meromorphic function on H with no zeros or poles on a contour $C\subset H$. Suppose that $f(\gamma z)=(cz+d)^k f(z)$. Then*

$$\int_C\frac{f'(z)}{f(z)}dz-\int_{\gamma C}\frac{f'(z)}{f(z)}dz=-k\int_C\frac{dz}{z+(d/c)}. \tag{2.24}$$

The required equality (2.23) follows immediately from the lemma, where we set $\gamma=S=\left(\begin{smallmatrix}0&-1\\1&0\end{smallmatrix}\right)$, note that $S(CD)=FE$, and compute that as $\varepsilon\to 0$

$$\frac{1}{2\pi i}\int_{CD}\frac{dz}{z}\to\int_{1/3}^{1/4}d\theta=-\frac{1}{12}\qquad(\text{where } z=e^{2\pi i\theta}).$$

Proof of the Lemma. Differentiating

$$f(\gamma z)=(cz+d)^k f(z), \tag{2.25}$$

we obtain

$$f'(\gamma z)\frac{d\gamma z}{dz}=(cz+d)^k f'(z)+kc(cz+d)^{k-1}f(z). \tag{2.26}$$

We now divide (2.26) by (2.25):

$$\frac{f'(\gamma z)}{f(\gamma z)}d\gamma z=\frac{f'(z)}{f(z)}dz+k\frac{cdz}{cz+d}.$$

Thus, the left side of (2.24) is equal to

$$\int_C\frac{f'(z)}{f(z)}dz-\frac{f'(\gamma z)}{f(\gamma z)}d\gamma z=-k\int_C\frac{cdz}{cz+d}.$$

This completes the proof of the lemma, and of Proposition 8. □

We now derive several very useful consequences of Proposition 8.

Proposition 9. *Let k be an even integer, $\Gamma=SL_2(\mathbb{Z})$.*

(a) *The only modular forms of weight 0 for Γ are constants, i.e., $M_0(\Gamma)=\mathbb{C}$.*
(b) *$M_k(\Gamma)=0$ if k is negative or $k=2$.*
(c) *$M_k(\Gamma)$ is one-dimensional, generated by E_k, if $k=4,6,8,10$ or 14; in other words, $M_k(\Gamma)=\mathbb{C}E_k$ for those values of k.*
(d) *$S_k(\Gamma)=0$ if $k<12$ or $k=14$; $S_{12}(\Gamma)=\mathbb{C}\Delta$; and for $k>14$ $S_k(\Gamma)=$*

$\Delta M_{k-12}(\Gamma)$ (*i.e., the cusp forms of weight k are obtained by multiplying modular forms of weight* $k - 12$ *by the function* $\Delta(z)$).

(e) $M_k(\Gamma) = S_k(\Gamma) \oplus \mathbb{C}E_k$ *for* $k > 2$.

PROOF. Note that for a modular *form* all terms on the left in (2.21) are nonnegative.

(a) Let $f \in M_0(\Gamma)$, and let c be any value taken by $f(z)$. Then $f(z) - c \in M_0(\Gamma)$ has a zero, i.e., one of the terms on the left in (2.21) is strictly positive. Since the right side is 0, this can only happen if $f(z) - c$ is the zero function.

(b) If $k < 0$ or $k = 2$, there is no way that the sum of nonnegative terms on the left in (2.21) could equal $k/12$.

(c) When $k = 4, 6, 8, 10$, or 14 we note that there is only one possible way of choosing the $v_P(f)$ so that (2.21) holds:
for $k = 4$, we must have $v_\omega(f) = 1$, all other $v_P(f) = 0$;
for $k = 6$, we must have $v_i(f) = 1$, all other $v_P(f) = 0$;
for $k = 8$, we must have $v_\omega(f) = 2$, all other $v_P(f) = 0$;
for $k = 10$, we must have $v_\omega(f) = v_i(f) = 1$, all other $v_P(f) = 0$;
for $k = 14$, we must have $v_\omega(f) = 2$, $v_i(f) = 1$, all other $v_P(f) = 0$.
Let $f_1(z)$, $f_2(z)$ be nonzero elements of $M_k(\Gamma)$. Since $f_1(z)$ and $f_2(z)$ have the same zeros, the weight zero modular function $f_1(z)/f_2(z)$ is actually a modular form. By part (a), f_1 and f_2 are proportional. Choosing $f_2(z) = E_k(z)$ gives part (c).

(d) For $f \in S_k(\Gamma)$ we have $v_\infty(f) > 0$, and all other terms on the left in (2.21) are nonnegative. Notice that when $k = 12$ and $f = \Delta$, (2.21) implies that the only zero of $\Delta(z)$ is at infinity. Hence, for any k and any $f \in S_k(\Gamma)$, the modular function f/Δ is actually a modular form, i.e., $f/\Delta \in M_{k-12}(\Gamma)$. This gives us all of the assertions in part (d).

(e) Since E_k does not vanish at infinity, given $f \in M_k(\Gamma)$ we can always subtract a suitable multiple of E_k so that the resulting $f - cE_k \in M_k(\Gamma)$ vanishes at infinity, i.e., $f - cE_k \in S_k(\Gamma)$. □

We now prove that any modular form for Γ is a polynomial in E_4, E_6 (see Problem 5 of §I.6 for a different proof of this fact).

Proposition 10. *Any* $f \in M_k(\Gamma)$ *can be written in the form*

$$f(z) = \sum_{4i+6j=k} c_{i,j} E_4(z)^i E_6(z)^j. \qquad (2.27)$$

PROOF. We use induction on k. For $k = 4, 6, 8, 10, 14$ we note that E_4, E_6, E_4^2, E_4E_6, $E_4^2E_6$, respectively, is an element of $M_k(\Gamma)$, and so, by Proposition 9(c), must span $M_k(\Gamma)$. Now suppose that $k = 12$ or $k > 14$. It is clearly possible to find i and j such that $4i + 6j = k$, in which case $E_4^iE_6^j \in M_k(\Gamma)$. Given $f \in M_k(\Gamma)$, by the same argument as in the proof of Proposition 9(e)

we can find $c \in \mathbb{C}$ such that $f - cE_4^i E_6^j \in S_k(\Gamma)$. By part (d) of Proposition 9, we can write f in the form

$$f = cE_4^i E_6^j + \Delta f_1 = cE_4^i E_6^j + \frac{(2\pi)^{12}}{1728}(E_4^3 - E_6^2)f_1,$$

where $f_1 \in M_{k-12}(\Gamma)$. Applying (2.27) to f_1 by the induction assumption (with k replaced by $k - 12$), we obtain the desired polynomial for f. □

The j-invariant. We now define a very important modular function of weight zero:

$$j(z) \underset{\text{def}}{=} \frac{1728 g_2(z)^3}{\Delta(z)} = 1728 \frac{E_4(z)^3}{E_4(z)^3 - E_6(z)^2} \qquad \text{(by (2.14)–(2.15))} \qquad (2.28)$$

Proposition 11. *The function j gives a bijection from $\Gamma \backslash \bar{H}$ (the fundamental domain with Γ-equivalent sides identified and the point at infinity included) and the Riemann sphere $\mathbb{P}_{\mathbb{C}}^1 = \mathbb{C} \cup \{\infty\}$.*

PROOF. In the proof of Proposition 9(d) we saw that $\Delta(z)$ has a simple zero at infinity and no other zero. Since g_2 does not vanish at infinity, this means that $j(z)$ has a simple pole at infinity and is holomorphic on H. For any $c \in \mathbb{C}$ the modular form $1728 g_2^3 - c\Delta \in M_{12}(\Gamma)$ must vanish at exactly one point $P \in \Gamma \backslash H$, because when $k = 12$ exactly one of the terms on the left in (2.21) is strictly positive. Dividing by Δ, we see that this means that $j(z) - c = 0$ for exactly one value of $z \in \Gamma \backslash H$. Thus, j takes ∞ to ∞ and on $\Gamma \backslash H$ is a bijection with $\mathbb{C}$. □

Proposition 12. *The modular functions of weight zero for Γ are precisely the rational functions of j.*

PROOF. A rational function of $j(z)$ is a modular function of weight zero (see Remark 3 at the beginning of this section). Conversely, suppose that $f(z)$ is a modular function of weight zero for Γ. If z_j are the poles of $f(z)$ in $\Gamma \backslash H$, counted with multiplicity, then $f(z) \cdot \Pi_j(j(z) - j(z_j))$ is a modular function of weight 0 with no poles in H, and it suffices to show that such a function is a rational function of j. So, without loss of generality we may assume that $f(z)$ has no poles in H. We can next multiply by a suitable power of Δ to cancel the pole of $f(z)$ at ∞. Thus, for some k we will have $\Delta(z)^k f(z) \in M_{12k}(\Gamma)$. By Proposition 10, we can write $f(z)$ as a linear combination of modular functions of the form $E_4^i E_6^j / \Delta^k$ (where $4i + 6j = 12k$), so it suffices to show that such a modular function is a rational expression in j. Since $4i + 6j$ is divisible by 12, we must have $i = 3i_0$ divisible by 3 and $j = 2j_0$ divisible by 2. But it is easy to check that E_4^3/Δ and E_6^2/Δ are each of the form $aj + b$, by (2.28); and $E_4^{3i_0} E_6^{2j_0}/\Delta^k$ is a product of such factors. This proves the proposition. □

The definition (2.28) is not the only way one could have defined the invariant. It is not hard to see that any ratio of two non-proportional modular forms of weight 12 would have satisfied Propositions 11 and 12. But $j(z)$ has the additional convenient properties that: its pole is at infinity, i.e., it is holomorphic on H; and its residue at the pole is 1, as we easily compute from (2.28) that the q-expansion of j starts out $\frac{1}{q} + \cdots$.

Recall that in Chapter I we used the Weierstrass $\wp$-function and its derivative to identify the analytic manifold $\mathbb{C}/L$ with an elliptic curve in $\mathbb{P}^2_{\mathbb{C}}$. Similarly, in our present context we have used the j-invariant to identify $\Gamma\backslash\bar{H}$ as an analytic manifold with the Riemann sphere $\mathbb{P}^1_{\mathbb{C}}$. Proposition 12 then amounts to saying that the only meromorphic functions on the Riemann sphere are the rational functions. Thus, Proposition 12 is analogous to Proposition 8 in §I.5, which characterized the field of elliptic functions as the rational functions of $x = \wp(z)$, $y = \wp'(z)$.

A loose end from Chapter I. In Chapter I we defined an elliptic curve over $\mathbb{C}$ to be a curve given by an equation of the form $y^2 = f(x)$, where $f(x)$ is a cubic with distinct roots. We then worked with curves whose equations were written in the form

$$y^2 = 4x^3 - g_2(L)x - g_3(L) \tag{2.29}$$

for some lattice L. It is not hard to see that a linear change of variables will bring an equation $y^2 = f(x)$, where $f(x)$ has distinct roots, into the form

$$y^2 = 4x^3 - Ax - B \quad \text{with} \quad A^3 - 27B^2 \neq 0. \tag{2.30}$$

But in order to write such a curve in the form (2.29) we need to show that a lattice L can always be found such that $g_2(L) = A$, $g_3(L) = B$. This was not proved in Chapter I, but it is easy to prove now using modular forms.

In what follows we shall write a lattice $L = \{m\omega_1 + n\omega_2\}$, where $z = \omega_1/\omega_2 \in H$, as follows: $L = \lambda L_z = \{m\lambda z + n\lambda\}$. Here $\lambda = \omega_2$. Thus, any lattice L is a complex multiple of (i.e., a rotation plus expansion of) a lattice of the form L_z.

Proposition 13. *For any A, $B \in \mathbb{C}$ such that $A^3 \neq 27B^2$ there exists $L = \lambda L_z$ such that*

$$g_2(L) = A; \tag{2.31}$$

$$g_3(L) = B. \tag{2.32}$$

Proof. It follows immediately from the definition of g_2 that $g_2(\lambda L_z) = \lambda^{-4} g_2(L_z)$, and similarly $g_3(\lambda L_z) = \lambda^{-6} g_3(L_z)$.

By (2.14), we can restate the proposition: there exist λ and z such that $E_4(z) = \lambda^4(3/4\pi^4)A$ and $E_6(z) = \lambda^6(27/8\pi^6)B$. Let $a = 3A/4\pi^4$, $b = 27B/8\pi^6$. Then the condition $A^3 \neq 27B^2$ becomes $a^3 \neq b^2$. Suppose we find a value of z such that $E_6(z)^2/E_4(z)^3 = b^2/a^3$. Then choose λ so that $E_4(z) = \lambda^4 a$,

in which case we must have

$$E_6(z)^2 = b^2 E_4(z)^3/a^3 = \lambda^{12} b^2, \quad \text{i.e.,} \quad E_6(z) = \pm\lambda^6 b.$$

If we have $+$ in the last equation, then our values of z and λ have the required properties (2.31)–(2.32). If we have $-$, then we need only replace λ by $i\lambda$. So it remains to find a value of z which gives the right ratio $E_6(z)^2/E_4(z)^3$. Now $E_6^2/E_4^3 = 1 - 1728/j$, by (2.28). Since $j(z)$ takes all finite values on H, and $b^2/a^3 \neq 1$, we can find a value of z such that $b^2/a^3 = 1 - 1728/j(z) = E_6(z)^2/E_4(z)^3$. This completes the proof. □

Thus, there was no loss of generality in Chapter I in taking our elliptic curves to be in the form (2.29) for some lattice L.

The Dedekind eta-function and the product formula for $\Delta(z)$. We conclude this section by proving the functional equation for the Dedekind η-function, from which Jacobi's product formula for $\Delta(z)$ will follow.

The function $\eta(z)$, $z \in H$, is defined by the product

$$\eta(z) = e^{2\pi i z/24} \prod_{n=1}^{\infty} (1 - e^{2\pi i n z}). \tag{2.33}$$

(In Problem 7 of §II.4 we gave another definition, and another proof of the functional equation. In the problems below we shall see the equivalence of the two definitions.)

Proposition 14. *Let $\sqrt{\ }$ denote the branch of the square root having nonnegative real part. Then*

$$\eta(-1/z) = \sqrt{z/i}\,\eta(z). \tag{2.34}$$

PROOF. The product (2.33) clearly converges to a nonzero value for any $z \in H$, and defines a holomorphic function on H. Suppose we show that the logarithmic derivatives of the left and right sides of (2.34) are equal. Then (2.34) must hold up to a multiplicative constant; but substituting $z = i$ shows that the constant must be 1.

Now the logarithmic derivative of (2.33) is

$$\frac{\eta'(z)}{\eta(z)} = \frac{2\pi i}{24}\left(1 - 24 \sum_{n=1}^{\infty} \frac{n e^{2\pi i n z}}{1 - e^{2\pi i n z}}\right).$$

If we expand each term in the sum as a geometric series in q^n ($q = e^{2\pi i z}$), and then collect terms with a given power of q, we find that

$$\frac{\eta'(z)}{\eta(z)} = \frac{2\pi i}{24}\left(1 - 24 \sum_{n=1}^{\infty} \sigma_1(n) q^n\right) = \frac{2\pi i}{24} E_2(z). \tag{2.35}$$

Meanwhile, the logarithmic derivative of the relation (2.34) that we want is:

$$\frac{\eta'(-1/z)}{\eta(-1/z)}z^{-2} = \frac{1}{2z} + \frac{\eta'(z)}{\eta(z)}. \tag{2.36}$$

Using (2.35), we reduce (2.36) to showing that

$$E_2(-1/z)z^{-2} = \frac{12}{2\pi iz} + E_2(z),$$

and this is precisely Proposition 7. □

Proposition 15.

$$(2\pi)^{-12}\Delta(z) = q\prod_{n=1}^{\infty}(1-q^n)^{24}, \qquad q = e^{2\pi iz}. \tag{2.37}$$

PROOF. Let $f(z)$ be the product on the right of (2.37). The function f is holomorphic on H, periodic of period 1, and vanishes at infinity. Moreover, $f(z)$ is the 24-th power of $\eta(z)$, by definition; and, raising both sides of (2.34) to the 24-th power, we find that $f(-1/z) = z^{12}f(z)$. Thus, $f(z)$ is a cusp form of weight 12 for Γ. By Proposition 9(d), $f(z)$ must be a constant multiple of $\Delta(z)$. Comparing the coefficient of q in their q-expansions, we conclude that $f(z)$ equals $(2\pi)^{-12}\Delta(z)$. □

Notice the role of the Eisenstein series $E_2(z)$ in proving the functional equation for $\eta(z)$, and then (2.37). The η-function turns up quite often in the study of modular forms. Some useful examples of modular forms, especially for congruence subgroups, can be built up from $\eta(z)$ and functions of the form $\eta(Mz)$.

The q-expansion in (2.37) is one of the famous series in number theory. Its coefficients are denoted $\tau(n)$ and called the Ramanujan function of n, since it was Ramanujan who proved or conjectured many of their properties:

$$\sum_{n=1}^{\infty}\tau(n)q^n \underset{\text{def}}{=} q\prod_{n=1}^{\infty}(1-q^n)^{24}.$$

Among the properties which will be shown later: (1) $\tau(n)$ is multiplicative ($\tau(nm) = \tau(n)\tau(m)$ if n and m are relatively prime); (2) $\tau(n) \equiv \sigma_{11}(n)$ (mod 691) (see Problem 4 below); (3) $\tau(n)/n^6$ is bounded. Ramanujan conjectured a stronger bound than (3), namely: $|\tau(n)| < n^{11/2}\sigma_0(n)$ (where $\sigma_0(n)$ is the number of divisors of n). The Ramanujan conjecture was finally proved ten years ago by Deligne as a consequence of his proof of the Weil conjectures. For more discussion of $\tau(n)$ and references for the proofs, see, for example, [Serre 1977] and [Katz 1976a].

PROBLEMS

1. Prove that for $k \geq 4$: $E_k(z) = \frac{1}{2}\sum_{\substack{m,n\in\mathbb{Z}\\(m,n)=1}}(mz+n)^{-k}$.

2. What is $E_2(i)$?

3. (a) Show that $E_4^2 = E_8$, $E_4E_6 = E_{10}$, and $E_6E_8 = E_{14}$.
   (b) Derive relations expressing σ_7 in terms of σ_3; σ_9 in terms of σ_3 and σ_5; and σ_{13} in terms of σ_5 and σ_7.

4. (a) Show that $E_{12} - E_6^2 = c\Delta$, with $c = (2\pi)^{-12}\cdot 2^6\cdot 3^5\cdot 7^2/691$.
   (b) Derive an expression for $\tau(n)$ in terms of σ_{11} and σ_5.
   (c) Prove that $\tau(n) \equiv \sigma_{11}(n) \pmod{691}$.

5. Prove that E_4 and E_6 are algebraically independent; in particular, the polynomial in Proposition 10 is unique.

6. (a) Prove the identity $\sum_{n=1}^{\infty} \frac{n^a x^n}{1-x^n} = \sum_{n=1}^{\infty} \sigma_a(n)x^n$.
   (b) Let $a \equiv 1 \pmod 4$ be an integer greater than 1. Show that $E_{a+1}(i) = 0$, and prove the following sequence of summation formulas:

$$\sum_{n=1}^{\infty} \frac{n^a}{e^{2\pi n} - 1} = \frac{1}{2(a+1)}B_{a+1} = \begin{cases} 1/504, & a = 5; \\ 1/264, & a = 9; \\ 1/24, & a = 13, \text{ etc.} \end{cases}$$

7. Let $f(z)$ be a modular form of weight k for Γ. Let

$$g(z) = \frac{1}{2\pi i} f'(z) - \frac{k}{12} E_2(z)f(z).$$

   Prove that $g(z)$ is a modular form of weight $k+2$ for Γ, and that it is a cusp form if and only if $f(z)$ is a cusp form.

8. (a) Prove that $E_6 = E_4E_2 - \frac{3}{2\pi i}E_4'$ and $E_8 = E_6E_2 - \frac{1}{\pi i}E_6'$.
   (b) Derive relations expressing σ_5 in terms of σ_1 and σ_3, and σ_7 in terms of σ_1 and σ_5.

9. Recall the following definitions and relations from Problems 4 and 7 in §II.4. Let χ be a nontrivial even primitive Dirichlet character mod N. Define

$$\theta(\chi, t) = \sum_{n=1}^{\infty} \chi(n)e^{-\pi t n^2}, \qquad \text{Re } t > 0.$$

   Then

$$\theta(\chi, t) = (N^2 t)^{-1/2} g(\chi)\theta(\bar{\chi}, 1/N^2 t).$$

   Now let χ be the character mod 12 such that $\chi(\pm 1) = 1$, $\chi(\pm 5) = -1$, and define $\tilde{\eta}(z) = \theta(\chi, -iz/12)$ for $z \in H$. Then $\tilde{\eta}(-1/z) = \sqrt{z/i}\,\tilde{\eta}(z)$.
   (a) Prove that $\tilde{\eta}(z+1) = e^{2\pi i/24}\tilde{\eta}(z)$, and that $\tilde{\eta}^{24} \in S_{12}(\Gamma)$.
   (b) Prove that $\tilde{\eta}(z) = \eta(z)$.
   (c) Write the equality in part (b) as an identity between formal power series in q. (*Note.* This identity is essentially Euler's "pentagonal number theorem." For a discussion of its combinatoric meaning and two more proofs of the identity, see [Andrews 1976, Corollaries 1.7 and 2.9].)

10. Find the j-invariant of the elliptic curve in Problem 3(b) of §I.2, which comes from the generalized congruent number problem. What is j in the classical case $\lambda = 1$?

Prove that there is no other value of $\lambda \in \mathbb{Q}$ for which the corresponding j is an integer. It is known that an elliptic curve with nonintegral j cannot have complex multiplication.

§3. Modular forms for congruence subgroups

Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma = SL_2(\mathbb{Z})$, let $f(z)$ be a function on $\bar{H} = H \cup \mathbb{Q} \cup \{\infty\}$ with values in $\mathbb{C} \cup \{\infty\}$, and let $k \in \mathbb{Z}$. We introduce the notation $f|[\gamma]_k$ to denote the function whose value at z is $(cz+d)^{-k}f((az+b)/(cz+d))$. We denote the value of $f|[\gamma]_k$ at z by $f(z)|[\gamma]_k$:

$$f(z)|[\gamma]_k \underset{\text{def}}{=} (cz+d)^{-k} f(\gamma z) \quad \text{for} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma. \tag{3.1}$$

More generally, let $GL_2^+(\mathbb{Q})$ denote the subgroup of $GL_2(\mathbb{Q})$ consisting of matrices with positive determinant. Then we define

$$f(z)|[\gamma]_k \underset{\text{def}}{=} (\det \gamma)^{k/2}(cz+d)^{-k} f(\gamma z) \quad \text{for} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2^+(\mathbb{Q}). \tag{3.2}$$

For example, if $\gamma = \left(\begin{smallmatrix} a & 0 \\ 0 & a \end{smallmatrix}\right)$ is a scalar matrix, we have $f|[\gamma]_k = f$ unless $a < 0$ and k is odd, in which case $f|[\gamma]_k = -f$.

Care should be taken with this notation. For example, by definition $f(2z)|[\gamma]_k$ means $f|[\gamma]_k$ evaluated at $2z$, i.e., $(2cz+d)^{-k}f((2az+b)/(2cz+d))$. This is *not* the same as $g(z)|[\gamma]_k$ for g the function defined by $g(z) = f(2z)$; namely, $g(z)|[\gamma]_k = (cz+d)^{-k}g(\gamma z) = (cz+d)^{-k}f(2(az+b)/(cz+d))$.

With this notation, any modular function of weight k for Γ satisfies $f|[\gamma]_k = f$ for all $\gamma \in \Gamma$. Some functions are invariant under $[\gamma]_k$ for other $\gamma \in GL_2^+(\mathbb{Q})$. For example, recall the theta-function defined in §4 of Chapter II: $\theta(t) = \sum_{n \in \mathbb{Z}} e^{-\pi t n^2}$ for $\operatorname{Re} t > 0$. We saw that it satisfies the functional equation $\theta(t) = t^{-1/2}\theta(1/t)$ (where $\sqrt{\ }$ is the branch such that $\sqrt{1} = 1$). We define a function of $z \in H$, also called the theta-function, by setting $\Theta(z) = \theta(-2iz)$, i.e.,

$$\Theta(z) = \sum_{n \in \mathbb{Z}} e^{2\pi i z n^2} = \sum_{n \in \mathbb{Z}} q^{n^2} \quad \text{for} \quad z \in H, \qquad q = e^{2\pi i z}. \tag{3.3}$$

Substituting $-2iz$ for t in the functional equation for θ, we have

$$\Theta(z) = (2z/i)^{-1/2}\Theta(-1/4z). \tag{3.4}$$

Squaring both sides and using the notation (3.2), we can write

$$\Theta^2|[\gamma]_1 = -i\Theta^2 \quad \text{for} \quad \gamma = \begin{pmatrix} 0 & -1 \\ 4 & 0 \end{pmatrix} \in GL_2^+(\mathbb{Q}). \tag{3.5}$$

Notice that

$$f|[\gamma_1\gamma_2]_k = (f|[\gamma_1]_k)|[\gamma_2]_k \quad \text{for} \quad \gamma_1, \gamma_2 \in GL_2^+(\mathbb{Q}). \tag{3.6}$$

To see this, write $(\det \gamma)^{k/2}(cz + d)^{-k}$ in the form $(d\gamma z/dz)^{k/2}$, and use the chain rule.

We are now ready to define modular functions, modular forms, and cusp-forms for a congruence subgroup $\Gamma' \subset \Gamma$. Let q_N denote $e^{2\pi i z/N}$.

Definition. Let $f(z)$ be a meromorphic function on H, and let $\Gamma' \subset \Gamma$ be a congruence subgroup of level N, i.e., $\Gamma' \supset \Gamma(N)$. Let $k \in \mathbb{Z}$. We call $f(z)$ a *modular function* of weight k for Γ' if

$$f|[\gamma]_k = f \quad \text{for all} \quad \gamma \in \Gamma', \tag{3.7}$$

and if, for any $\gamma_0 \in \Gamma = SL_2(\mathbb{Z})$,

$$f(z)|[\gamma_0]_k \quad \text{has the form} \quad \sum a_n q_N^n \quad \text{with} \quad a_n = 0 \quad \text{for} \quad n \ll 0. \tag{3.8}$$

We call such an $f(z)$ a modular *form* of weight k for Γ' if it is holomorphic on H and if for all $\gamma_0 \in \Gamma$ we have $a_n = 0$ for all $n < 0$ in (3.8). We call a modular form a *cusp-form* if in addition $a_0 = 0$ in (3.8) for all $\gamma_0 \in \Gamma$.

Thus, as in the case $\Gamma' = \Gamma$ treated in §2, a modular "function" is allowed to have poles of finite order, a "form" must be holomorphic at all points including the cusps, and a "cusp-form" must vanish at *all* cusps. This interpretation of (3.8) as a condition "at the cusps" will be explained below.

The first condition (3.7) is the obvious analog of the first condition (2.1) for modular functions for Γ. The second condition (3.8) is called "meromorphicity" at the cusps ("holomorphicity" if $a_n = 0$ for $n < 0$, "vanishing" if $a_n = 0$ for $n \le 0$). We now explain this further.

Let $g = f|[\gamma_0]_k$ for some fixed $\gamma_0 \in GL_2^+(\mathbb{Q})$. If f is invariant under Γ', i.e., if $f|[\gamma]_k = f$ for $\gamma \in \Gamma'$, then it follows from (3.6) that g is invariant under the group $\gamma_0^{-1}\Gamma'\gamma_0$: for all $\gamma_0^{-1}\gamma\gamma_0 \in \gamma_0^{-1}\Gamma'\gamma_0$ we have $g|[\gamma_0^{-1}\gamma\gamma_0]_k = (f|[\gamma_0]_k)|[\gamma_0^{-1}\gamma\gamma_0]_k = f|[\gamma\gamma_0]_k = (f|[\gamma]_k)|[\gamma_0]_k = f|[\gamma_0]_k = g$. In particular, if $\gamma_0 \in \Gamma$ and $\Gamma' \supset \Gamma(N)$, then $\gamma_0^{-1}\Gamma'\gamma_0$ also contains $\Gamma(N)$ (since $\Gamma(N)$ is normal in Γ), and so $g = f|[\gamma_0]_k$ is invariant under $\Gamma(N)$. Because $T^N = \left(\begin{smallmatrix}1 & N\\ 0 & 1\end{smallmatrix}\right) \in \Gamma(N)$, we have $g(z + N) = g(z)$, and so $g = f|[\gamma_0]_k$ has a Fourier series expansion in powers of $q_N = e^{2\pi i z/N}$. The content of condition (3.8) is that this expansion has only finitely many negative powers of q_N (no negative powers for holomorphicity, only positive powers for vanishing).

It may happen that $g = f|[\gamma_0]_k$ is invariant under a smaller translation T^h, where $h|N$, i.e., $g(z + h) = g(z)$. In that case the only powers of q_N that appear in the Fourier series are powers of $q_h = q_N^{N/h}$. For example, if $\gamma_0 = I$ and $\Gamma' = \Gamma_0(N)$, then $T \in \Gamma'$ and $g(z + 1) = g(z)$. In that case $g = f$ has an expansion in powers of $q = q_1 = q_N^N$. On the other hand, for $\Gamma' = \Gamma_0(N)$ and $\gamma_0 = S = \left(\begin{smallmatrix}0 & -1\\ 1 & 0\end{smallmatrix}\right)$ we have $\gamma_0^{-1}\Gamma'\gamma_0 = \Gamma^0(N) = \{\left(\begin{smallmatrix}* & 0\\ * & *\end{smallmatrix}\right) \bmod N\}$; thus, if f is a modular function for $\Gamma_0(N)$, in general we only have $g(z + h) = g(z)$ for $h = N$, where $g = f|[S]_k$.

Any cusp $s \in \mathbb{Q} \cup \{\infty\}$ can be written in the form $s = \alpha^{-1}\infty$ for some $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ (by writing $s = -d/c$ in lowest terms and finding any solution a and b to the equation $ad - bc = 1$). If we set $\gamma_0 = \alpha^{-1}$ in (3.8), the behavior of $g = f|[\alpha^{-1}]_k$ near ∞ (i.e., near $q_N = 0$) is a reflection of the behavior of f near s, since $g(z) = (-cz + a)^{-k} f((dz - b)/(-cz + a))$.

Proposition 16. *The condition* (3.8) *depends only on the* Γ'*-equivalence class of* $s = \gamma_0\infty$. *More precisely, if* $\gamma_1\infty = \gamma'\gamma_2\infty$ *for some* $\gamma' \in \Gamma'$, *then the smallest power of* q_N *that occurs in the Fourier expansion of* $f|[\gamma_1]_k$ *and* $f|[\gamma_2]_k$ *is the same. Moreover, if this smallest power is the constant term, then the value at* $q_N = 0$ *is the same for* $f|[\gamma_1]_k$ *and* $f|[\gamma_2]_k$ *if* k *is even; if* k *is odd, this value may at most change sign.*

PROOF. If $\gamma_1\infty = \gamma'\gamma_2\infty$, then the element $\gamma_1^{-1}\gamma'\gamma_2 \in \Gamma$ keeps ∞ fixed, in which case it must be of the form $\pm T^j$. (Note that $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)\infty = \infty$ is equivalent to $c = 0$, and the elements of Γ with $c = 0$ are $\pm T^j$.) Thus, we have $\gamma_1^{-1}\gamma'\gamma_2 = \pm T^j$, so that $\gamma_2 = \pm\gamma'^{-1}\gamma_1 T^j$. Let $g(z) = f(z)|[\gamma_1]_k = \Sigma\, a_n q_N^n$. Since $f|[-I]_k = (-1)^k f$, and $f|[\gamma'^{-1}]_k = f$ (because $\gamma' \in \Gamma'$), it follows that $f|[\gamma_2]_k = (\pm 1)^k (f|[\gamma_1]_k)|[T^j]_k = (\pm 1)^k g|[T^j]_k$. Thus,

$$f(z)|[\gamma_2]_k = (\pm 1)^k g(z + j) = (\pm 1)^k \sum a_n e^{2\pi i n j/N} q_N^n.$$

In other words, the q_N-expansion coefficients corresponding to γ_2 differ from those corresponding to γ_1 only by roots of unity. The proposition now follows immediately. □

If f is a meromorphic function on H which is invariant under $[\gamma']_k$ for $\gamma' \in \Gamma'$, and if $s \in \mathbb{Q} \cup \{\infty\}$ with $s = \gamma_0\infty$, $\gamma_0 \in \Gamma$, then we say that f is meromorphic (is holomorphic, vanishes) at the cusp s if $f|[\gamma_0]_k$ has a Fourier expansion with only finitely many negative terms (respectively, with no negative terms, with no negative terms or constant term). Proposition 16 says that meromorphicity, holomorphicity, vanishing at s does not depend on the choice of γ_0 for which $s = \gamma_0\infty$, and in fact only depends on the Γ'-equivalence class of s.

Thus, the condition (3.8) is really a set of conditions, one corresponding to each cusp s of Γ'. (Recall that "cusp of Γ'" means "Γ'-equivalence class of cusps".) For example, if $\Gamma' = \Gamma_0(p)$ for p a prime, we saw (Problem 18 in §III.1) that there are only two cusps ∞, 0. Thus, the condition (3.8) amounts to the two conditions

$$f(z) = \sum a_n q^n, \qquad q = e^{2\pi i z}, \quad \text{with} \quad a_n = 0 \quad \text{for} \quad n \ll 0; \tag{3.9}$$

$$z^{-k} f(-1/z) = \sum b_n q_p^n, \qquad q_p = e^{2\pi i z/p}, \quad \text{with} \quad b_n = 0 \quad \text{for} \quad n \ll 0. \tag{3.10}$$

We call the Fourier series in (3.9) the q-expansion of f at ∞, and we call (3.10) the q_p-expansion of f at the cusp 0. If f is holomorphic, we write $f(\infty)$ for a_0 and $f(0)$ for b_0. Note that $f(0)$ is *not* the limit of $f(z)$ as $z \to 0$.

We let $M_k(\Gamma')$ and $S_k(\Gamma')$ denote the set of modular forms of weight k for Γ' and the set of cusp-forms of weight k for Γ', respectively. As in the case $\Gamma' = \Gamma$ treated in the last section, it is easy to see that these are $\mathbb{C}$-vector spaces, that $f \in M_{k_1}(\Gamma')$ and $g \in M_{k_2}(\Gamma')$ implies $fg \in M_{k_1+k_2}(\Gamma')$, and that the vector space of weight zero modular functions for Γ' is a field. Also note that if $-I \in \Gamma'$, then there are no nonzero modular functions for Γ' of odd weight k, since then $f|[-I]_k = -f$.

It is immediate from the definition that if $\Gamma'' \subset \Gamma'$, then a modular function/modular form/cusp-form for Γ' is also a modular function/modular form/cusp-form for Γ''.

There are more interesting ways to get modular forms for a congruence subgroup Γ'' from forms for another subgroup Γ'. For example, if $f(z) = \Sigma\, a_n q^n \in M_k(\Gamma)$, then $f(Nz) = \Sigma\, a_n q^{Nn}$ and $f_\chi(z) = \Sigma\, a_n \chi(n) q^n$ (the "twist" of f by a Dirichlet character χ) turn out to be modular forms, although for a smaller congruence subgroup than Γ. The next proposition gives two important classes of constructions of this type. In part (b) of the proposition we use the notation $M_k(N, \chi)$ with χ a Dirichlet character mod N to denote the subspace of $M_k(\Gamma_1(N))$ (see (1.5)) consisting of $f(z)$ for which $f|[\gamma]_k = \chi(d) f$ whenever $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$. In particular, for χ the trivial character $M_k(N, \chi_{\text{triv}}) = M_k(\Gamma_0(N))$.

Proposition 17. (a) *Let Γ' be a congruence subgroup of Γ, let $\alpha \in GL_2^+(\mathbb{Q})$, and set $\Gamma'' = \alpha^{-1}\Gamma'\alpha \cap \Gamma$. Then Γ'' is a congruence subgroup of Γ, and the map $f \mapsto f|[\alpha]_k$ takes $M_k(\Gamma')$ to $M_k(\Gamma'')$, and takes $S_k(\Gamma')$ to $S_k(\Gamma'')$. In particular, if $f \in M_k(\Gamma)$ and $g(z) = f(Nz)$, then $g \in M_k(\Gamma_0(N))$ and one has $g(\infty) = f(\infty)$, $g(0) = N^{-k} f(0)$.*

(b) *Let χ be a Dirichlet character modulo M, and let χ_1 be a primitive Dirichlet character modulo N. If $f(z) = \Sigma_{n=0}^{\infty} a_n q^n \in M_k(M, \chi)$ and $f_{\chi_1}(z) \underset{\text{def}}{=} \Sigma_{n=0}^{\infty} a_n \chi_1(n) q^n$, then $f_{\chi_1} \in M_k(MN^2, \chi\chi_1^2)$. If f is a cusp form, then so is f_{χ_1}. In particular, if $f \in M_k(\Gamma_0(M))$ and χ_1 is quadratic (i.e., takes values ± 1), then $f_{\chi_1} \in M_k(\Gamma_0(MN^2))$.*

Proof. (a) We need two lemmas.

Lemma 1. *Let $\alpha \in GL_2^+(\mathbb{Q})$ have integer entries, and let $D = \det \alpha$. If $\Gamma' \supset \Gamma(N)$, then $\alpha^{-1}\Gamma'\alpha \supset \Gamma(ND)$.*

Proof of Lemma 1. Suppose $\gamma \in \Gamma(ND)$, i.e., $\gamma = 1 + ND\beta$ for some 2×2-matrix β with integer entries and $\det \gamma = 1$. We must show that $\gamma \in \alpha^{-1}\Gamma'\alpha$, i.e., that $\Gamma' \ni \alpha\gamma\alpha^{-1} = \alpha(1 + ND\beta)\alpha^{-1} = 1 + ND\alpha\beta\alpha^{-1}$. But $\alpha' = D\alpha^{-1}$ is an integer matrix. Since $\det \alpha\gamma\alpha^{-1} = \det \gamma = 1$ and $\alpha\gamma\alpha^{-1} = 1 + N\alpha\beta\alpha'$, we have $\alpha\gamma\alpha^{-1} \in \Gamma(N) \subset \Gamma'$, as claimed. □

Lemma 2. *Suppose that $f(z)$ has the property* (3.8) *for all $\gamma_0 \in \Gamma$, i.e., $f(z)|[\gamma_0]_k = \Sigma_{n=n_0}^{\infty} a_n q_N^n$ (where $n_0 = 0$ if $f(z)$ is holomorphic at the cusps, $n_0 = 1$ if $f(z)$ vanishes at the cusps). Then $f(z)$ has the same property for all $\alpha \in GL_2^+(\mathbb{Q})$, i.e., $f(z)|[\alpha]_k = \Sigma_{n=an_0}^{\infty} b_n q_{ND}^n$ (for some positive integers a and D which depend on α).*

PROOF OF LEMMA 2. Since α can be multiplied by a positive scalar without affecting $[\alpha]_k$, without loss of generality we may suppose that α has integer entries. It is an easy exercise in linear algebra to show that there exists $\gamma_0 \in \Gamma = SL_2(\mathbb{Z})$ such that $\gamma_0^{-1}\alpha = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$, where a and d are positive integers. Then

$$\begin{aligned} f(z)|[\alpha]_k &= (f(z)|[\gamma_0]_k)|\left[\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}\right]_k \\ &= (ad)^{k/2}d^{-k}\sum_{n=n_0}^{\infty} a_n e^{2\pi i n((az+b)/d)/N} \\ &= (a/d)^{k/2}\sum_{n=n_0}^{\infty} a_n e^{2\pi i b n/dN} q_{Nd}^{an} \\ &= \sum_{n=an_0}^{\infty} b_n q_{Nd}^{n}, \end{aligned}$$

where

$$b_n = \begin{cases} 0 & \text{if } a \nmid n; \\ (a/d)^{k/2} e^{2\pi i n b/adN} a_{n/a} & \text{if } a | n. \end{cases}$$

This proves the lemma. □

We now turn to the proof of the proposition. The first assertion in part (c) follows from Lemma 1. Now suppose that $f \in M_k(\Gamma')$. Then for $\alpha^{-1}\gamma'\alpha \in \Gamma''$ (where $\gamma' \in \Gamma'$) we have $(f|[\alpha]_k)|[\alpha^{-1}\gamma'\alpha]_k = (f|[\gamma']_k)|[\alpha]_k = f|[\alpha]_k$; in addition, for $\gamma_0 \in \Gamma$ we have $(f|[\alpha]_k)|[\gamma_0]_k = f|[\alpha\gamma_0]_k$, and the condition (3.8) holds for $f|[\alpha]_k$, by Lemma 2.

Thus, $f|[\alpha]_k \in M_k(\Gamma'')$. If f vanishes at all of the cusps, then so does $f|[\alpha]_k$, by Lemma 2. To obtain the last assertion in Proposition 17(a), we write $\alpha = \left(\begin{smallmatrix} N & 0 \\ 0 & 1 \end{smallmatrix}\right)$, $g = N^{-k/2} f|[\alpha]_k$, and note that $\alpha^{-1}\Gamma\alpha \cap \Gamma = \Gamma_0(N)$. The values at the two cusps ∞ and 0 come from the above formula for b_n with $n = 0$ and α replaced by $\left(\begin{smallmatrix} N & 0 \\ 0 & 1 \end{smallmatrix}\right)$ at the cusp ∞ and by $\left(\begin{smallmatrix} N & 0 \\ 0 & 1 \end{smallmatrix}\right)S = S\left(\begin{smallmatrix} 1 & 0 \\ 0 & N \end{smallmatrix}\right)$ at the cusp 0. This completes the proof of Proposition 17(a).

(b) Let $\xi = e^{2\pi i/N}$, and let $g = \sum_{j=0}^{N-1} \chi_1(j)\xi^j$ be the Gauss sum. Then

$$\begin{aligned} f_{\chi_1}(z) &= \sum_{l=0}^{N-1} \chi_1(l) \sum_{n=0}^{\infty} \left(\frac{1}{N}\sum_{\nu=0}^{N-1} \xi^{(l-n)\nu}\right) a_n q^n \\ &= \frac{1}{N}\sum_{l,\nu=0}^{N-1} \bar{\chi}_1(\nu)\chi_1(l\nu)\xi^{l\nu} \sum_{n=0}^{\infty} a_n e^{2\pi i n(z-\nu/N)} \\ &= \frac{g}{N}\sum_{\nu=0}^{N-1} \bar{\chi}_1(\nu) f(z-\nu/N). \end{aligned}$$

Now let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(MN^2)$. We want to examine $f_{\chi_1}(\gamma z)$. Let γ_ν denote $\left(\begin{smallmatrix} 1 & -\nu/N \\ 0 & 1 \end{smallmatrix}\right)$. Then for each ν and ν', $0 \le \nu, \nu' < N$, we compute

$$\gamma_\nu \gamma \gamma_{\nu'}^{-1} = \begin{pmatrix} a - c\nu/N & b + (\nu' a - \nu d)/N - c\nu\nu'/N^2 \\ c & d + c\nu'/N \end{pmatrix}.$$

invariant under $[\gamma]_k$ for $\gamma \in \Gamma_1(N)$. It remains to check the holomorphicity condition at the cusps. But $[\gamma_0]_k$ permutes the $G_k^{\underline{a} \bmod N}$ for $\gamma_0 \in \Gamma$, by (3.13). Hence it suffices to show that each $G_k^{\underline{a}}$ is finite at infinity. But

$$\lim_{z \to i\infty} G_k^{\underline{a}}(z) = \sum_{\underline{m} \equiv \underline{a} \bmod N, m_1 = 0} m_2^{-k} = \begin{cases} 0 & \text{if } a_1 \neq 0; \\ \sum_{n \equiv a_2 \bmod N} n^{-k} & \text{if } a_1 = 0. \end{cases}$$

The sum Σn^{-k} over $n \equiv a_2 \bmod N$ converges because $k \geq 3$. It is essentially a "partial" zeta-function. More precisely,

$$G_k^{(0,a_2)}(\infty) = \zeta^{a_2}(k) + (-1)^k \zeta^{-a_2}(k), \quad \text{where } \zeta^a(k) \underset{\text{def}}{=} \sum_{\substack{n \geq 1 \\ n \equiv a \bmod N}} n^{-k}. \tag{3.14}$$

This completes the proof of the proposition. □

As a special case of (3.13), if we set $\gamma = -I$ we have

$$G_k^{-\underline{a}} = (-1)^k G_k^{\underline{a}}. \tag{3.15}$$

This can also be seen directly from the definition (3.12). Thus, for example, $G_k^{\underline{a}} = 0$ if k is odd and $2\underline{a} \equiv 0 \bmod N$.

It is now not hard to construct modular forms for any congruence subgroup Γ', $\Gamma \supset \Gamma' \supset \Gamma(N)$, out of the Eisenstein series $G_k^{\underline{a} \bmod N}$. For fixed $\underline{a}$, the elements $\gamma \in \Gamma'$ permute the Eisenstein series $G_k^{\underline{a}'}$, where $\underline{a}'$ ranges over the orbit of $\underline{a}$ under the action of Γ', i.e.,

$$\underline{a}' \in \underline{a}\Gamma' \underset{\text{def}}{=} \{\underline{a}\gamma | \gamma \in \Gamma'\}.$$

Let $r = \#(\underline{a}\Gamma')$ be the number of elements in the orbit. If $F(X_1, \ldots, X_r)$ is any homogeneous symmetric polynomial in r variables with total degree d, and we set $X_j = G_k^{\underline{a}\gamma_j}$ (where $\underline{a}\gamma_j$ runs through the orbit $\underline{a}\Gamma'$), then $F(G_k^{\underline{a}\gamma_1}, \ldots, G_k^{\underline{a}\gamma_r})$ is easily seen to be a modular form of weight kd for Γ'. For example, taking F to be $X_1 + \cdots + X_r$ or $X_1 \ldots X_r$, we have

$$\begin{gathered} \sum_{a_2 \in (\mathbb{Z}/N\mathbb{Z})^*} G_k^{(0,a_2) \bmod N}(z) \in M_k(\Gamma_0(N)); \\ \prod_{a_2 \in (\mathbb{Z}/N\mathbb{Z})^*} G_k^{(0,a_2) \bmod N}(z) \in M_{\phi(N)k}(\Gamma_0(N)). \end{gathered} \tag{3.16}$$

We now compute the q_N-expansion for $G_k^{\underline{a}}$. We shall be especially interested in the cases when $a_1 = 0$ or $a_2 = 0$. Recall the formula used in deriving the q-expansion for G_k in the last section (see the proof of Proposition 6):

$$\sum_{n \in \mathbb{Z}} \frac{1}{(z+n)^k} = (-1)^{k-1} 2\zeta(k) \frac{k}{B_k} \sum_{j=1}^{\infty} j^{k-1} e^{2\pi i j z} \quad \text{for} \quad k \geq 2, \qquad z \in H. \tag{3.17}$$

Let

Since $g|[T]_8 = g$, and $[\frac{1}{2}I]_k$ is always trivial, it follows that g is invariant under $\frac{1}{2}\alpha T\alpha$, as desired. □

Eisenstein series. Let N be a positive integer. Let $\underline{a} = (a_1, a_2)$ be a pair of integers modulo N. We shall either consider the a_i as elements of $\mathbb{Z}/N\mathbb{Z}$ or as integers in the range $0 \le a_i < N$. Let k be an integer at least 3. Let $\underline{m} = (m_1, m_2)$ denote a pair of integers. We shall think of $\underline{a}$ and $\underline{m}$ as row vectors. For $z \in H$ we define the "level N Eisenstein series" (corresponding to $\underline{a}$ and k) as follows:

$$G_k^{\underline{a}}(z) = G_k^{\underline{a} \bmod N}(z) \underset{\text{def}}{=} \sum_{\substack{\underline{m} \in \mathbb{Z}^2 \\ \underline{m} \equiv \underline{a} \bmod N}} \frac{1}{(m_1 z + m_2)^k}. \tag{3.12}$$

If $\underline{a} = (0, 0)$, we delete the pair $\underline{m} = (0, 0)$ in the sum (3.12). But there is no point in considering the case when $\underline{a} = \underline{0}$, since, setting $m_1 = Nm$, $m_2 = Nn$, we have

$$G_k^{\underline{0}}(z) = N^{-k} \sum_{m,n \in \mathbb{Z}}' (mz + n)^{-k} = N^{-k} G_k(z),$$

which is the Eisenstein series for Γ which we already studied in §2. In what follows we shall suppose that $\underline{a} \ne (0, 0)$.

Notice that we are allowing $k \ge 3$ to be either odd or even.

Proposition 21. *$G_k^{\underline{a} \bmod N} \in M_k(\Gamma(N))$, and $G_k^{(0,a_2) \bmod N} \in M_k(\Gamma_1(N))$.*

PROOF. First, the series (3.12) is absolutely and uniformly convergent for z in any compact subset of H, because $k \ge 3$ (see, for example, Problem 3 in §I.5). Hence $G_k^{\underline{a}}(z)$ is holomorphic in H. Now let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. Then

$$G_k^{\underline{a}}(z)|[\gamma]_k = (cz + d)^{-k} \sum_{\underline{m} \equiv \underline{a} \bmod N} \frac{1}{\left(m_1 \dfrac{az + b}{cz + d} + m_2\right)^k}$$

$$= \sum_{\underline{m} \equiv \underline{a} \bmod N} \frac{1}{((m_1 a + m_2 c)z + (m_1 b + m_2 d))^k}.$$

Let $\underline{m}' = (m_1 a + m_2 c, m_1 b + m_2 d) = (m_1 \; m_2)\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \underline{m}\gamma$. Note that modulo N we have $\underline{m}' \equiv \underline{a}\gamma$. Let $\underline{a}' = \underline{a}\gamma$ (reduced modulo N). Then the maps $\underline{m} \mapsto \underline{m}' = \underline{m}\gamma$, and $\underline{m}' \mapsto \underline{m} = \underline{m}'\gamma^{-1}$ give a one-to-one correspondence between pairs $\underline{m} \in \mathbb{Z}^2$ with $\underline{m} \equiv \underline{a} \bmod N$ and pairs $\underline{m}' \in \mathbb{Z}^2$ with $\underline{m}' \equiv \underline{a}' \bmod N$. This means that the last sum above is equal to $\sum_{\underline{m}' \equiv \underline{a}' \bmod N} (m_1' z + m_2')^{-k} = G_k^{\underline{a}'}(z)$. Thus,

$$G_k^{\underline{a} \bmod N}|[\gamma]_k = G_k^{\underline{a}\gamma \bmod N} \quad \text{for} \quad \gamma \in \Gamma. \tag{3.13}$$

If $\gamma \in \Gamma(N)$, then by definition $\gamma \equiv I \bmod N$, and so $\underline{a}\gamma \equiv \underline{a} \bmod N$. Thus, (3.13) shows that $G_k^{\underline{a}}$ is invariant under $[\gamma]_k$ for $\gamma \in \Gamma(N)$. Similarly, if $\gamma \in \Gamma_1(N)$ and $a_1 = 0$, we have $(0, a_2)\gamma \equiv (0, a_2) \bmod N$, and so $G_k^{(0,a_2) \bmod N}$ is

cusp-form of weight k for $\Gamma_0(N)$ must be a multiple of $(\eta(z)\eta(Nz))^k$. Here $\eta(z) = e^{2\pi iz/24}\prod_n(1-q^n)$, $q = e^{2\pi iz}$, as in §2.

Proposition 19. *Let $f(z)$ be a nonzero element of $S_k(\Gamma_0(N))$, where $N = 2$, 3, 5, or 11 and $k = 8$, 6, 4, or 2, respectively, so that $k(N+1) = 24$. Then $f(z)$ is a constant multiple of $g(z) \underset{\text{def}}{=} (\eta(z)\eta(Nz))^k$.*

Proof. By Proposition 17(a), $g(z)^{N+1} = \Delta(z)\Delta(Nz)$ is an element of $S_{24}(\Gamma_0(N))$. In addition, $g(z)^{N+1}$ is nonzero on H, since $\Delta(z) \neq 0$ on H. At infinity, $g(z)^{N+1} = q^{N+1}\prod(1-q^n)^{24}(1-q^{Nn})^{24}$ has a zero of order $N+1$ in its q-expansion. At the cusp zero, we write

$$g(z)^{N+1}|[S]_{24} = z^{-24}\Delta(-1/z)\Delta(-N/z) = z^{-24}z^{12}\Delta(z)(z/N)^{12}\Delta(z/N)$$
$$= N^{-12}q_N^{N+1}\prod(1-q_N^n)^{24}(1-q_N^{Nn})^{24},$$

which has a zero of order $N+1$ in its q_N-expansion. On the other hand, since $f \in S_k(\Gamma_0(N))$, its q-expansion at ∞ is divisible by q, and the q_N-expansion of $f|[S]_k$ is divisible by q_N. Now $(f/g)^{N+1}$, as a ratio of two elements of $S_{24}(\Gamma_0(N))$, is a modular function of weight zero for $\Gamma_0(N)$. Since $g(z) \neq 0$ on H, this ratio is holomorphic on H. Moreover, the q-expansion of f^{N+1} is divisible at least by q^{N+1}, and at zero the q_N-expansion is divisible at least by q_N^{N+1}. Hence, the q^{N+1} in g^{N+1} and the q_N^{N+1} in $g^{N+1}|[S]_{24}$ are canceled, and the ratio is holomorphic at the cusps, i.e., $(f/g)^{N+1} \in M_0(\Gamma_0(N))$. By Proposition 18, $(f/g)^{N+1}$ is a constant, and hence f/g is also a constant. This completes the proof. □

Notice that we did not actually prove that $g(z) = (\eta(z)\eta(Nz))^k$ is in $S_k(\Gamma_0(N))$, unless we can be assured that there exists a nonzero element $f \in S_k(\Gamma_0(N))$. The same proof, for example, would tell us that any nonzero $f \in S_3(\Gamma_0(7))$ must be a constant multiple of $(\eta(z)\eta(7z))^3$; but $S_3(\Gamma_0(7)) = 0$, since 3 is odd and $-I \in \Gamma_0(7)$. However, for the values of N and k in Proposition 19 it can be shown that $\dim S_k(\Gamma_0(N)) = 1$ (see Theorem 2.24 and Proposition 1.43 in Chapter 2 of [Shimura 1971]). Thus, Proposition 19 is not vacuous. For $N = 2$, $k = 8$ we can see this directly, since we know that T and $ST^2S = \left(\begin{smallmatrix}-1 & 0\\ 2 & -1\end{smallmatrix}\right)$ generate $\overline{\Gamma}_0(2)$ (see Problem 13(b) of §III.1).

Proposition 20. $(\eta(z)\eta(2z))^8 \in S_8(\Gamma_0(2))$.

Proof. Clearly, $g(z) = (\eta(z)\eta(2z))^8$ is holomorphic on H. Its q-expansion at ∞ is: $(e^{2\pi iz/24+2\pi i2z/24})^8\prod(1-q^n)^8(1-q^{2n})^8 = q\prod(1-q^n)^8(1-q^{2n})^8$. Using the relation $\eta(-1/z) = \sqrt{z/i}\,\eta(z)$, we easily see that $g(z)$ vanishes at the cusp 0 as well. It remains to show invariance under $[\left(\begin{smallmatrix}-1 & 0\\ 2 & -1\end{smallmatrix}\right)]_8$. Set $\alpha = \left(\begin{smallmatrix}0 & -1\\ 2 & 0\end{smallmatrix}\right)$. Then $\left(\begin{smallmatrix}-1 & 0\\ 2 & -1\end{smallmatrix}\right) = \frac{1}{2}\alpha T\alpha$, and

$$g(z)|[\alpha]_8 = 2^4(2z)^{-8}(\eta(-1/2z)\eta(-1/z))^8 = (2z^2)^{-4}(\sqrt{2z/i}\,\eta(2z)\sqrt{z/i}\,\eta(z))^8$$
$$= (\eta(z)\eta(2z))^8 = g(z).$$

If v' is chosen for each v so that $v'a \equiv vd \bmod N$ (such a choice is unique, because a and d are prime to N), then we have $\gamma_v\gamma\gamma_{v'}^{-1} \in \Gamma_0(M)$, and so

$$\begin{aligned} f_{\chi_1}(\gamma z) &= \frac{g}{N}\sum_{v=0}^{N-1} \bar{\chi}_1(v) f(\gamma_v\gamma\gamma_{v'}^{-1}\gamma_{v'}z) \\ &= \frac{g}{N}\sum_{v=0}^{N-1} \bar{\chi}_1(v)\chi(d)(c\gamma_{v'}z + d + cv'/N)^k f(\gamma_{v'}z) \\ &= \chi(d)(cz+d)^k \frac{g}{N}\sum_{v=0}^{N-1} \bar{\chi}_1(v) f(z - v'/N). \end{aligned}$$

But $\bar{\chi}_1(v) = \bar{\chi}_1(v')\bar{\chi}_1(a)/\bar{\chi}_1(d) = \chi_1(d)^2\bar{\chi}_1(v')$. Thus,

$$f_{\chi_1}(\gamma z) = \chi\chi_1^2(d)(cz+d)^k \frac{g}{N}\sum_{v=0}^{N-1} \bar{\chi}_1(v') f(z - v'/N) = \chi\chi_1^2(d)(cz+d)^k f_{\chi_1}(z),$$

and so f_{χ_1} has the right transformation formula to be in $M_k(MN^2, \chi\chi_1^2)$. The cusp conditions are verified by the same method as in the proof of part (a) of the proposition. Namely, for all $\gamma \in \Gamma$, $f_{\chi_1}(z)|[\gamma]_k$ is a linear combination of $f(z)|[\gamma_v\gamma]_k$, and so the cusp conditions follow from Lemma 2. □

The next proposition generalizes Proposition 9(a) in the last section. Like Proposition 9(a), it is useful in proving equality of two modular forms from information about their zeros.

Proposition 18. $M_0(\Gamma') = \mathbb{C}$ *for any congruence subgroup* $\Gamma' \subset \Gamma$. *That is, there are no non-constant modular forms of weight zero.*

PROOF. Let $f \in M_0(\Gamma')$, and let $a = f(z_0)$ for some fixed $z_0 \in H$. Let $\Gamma = \bigcup \alpha_j\Gamma'$ be a disjoint union of cosets, and consider $g \underset{\text{def}}{=} \Pi\,(f|[\alpha_j^{-1}]_0 - a)$, i.e.,

$$g(z) = \prod (f(\alpha_j^{-1}z) - a). \tag{3.11}$$

Then $g(z)$ is holomorphic on H, and it satisfies (3.8), because f does. Moreover, given $\gamma^{-1} \in \Gamma$ we have $g|[\gamma^{-1}]_0 = \Pi\,(f|[(\gamma\alpha_j)^{-1}]_0 - a)$. But since $f|[\alpha^{-1}]_0$ does not change if α is replaced by another element $\alpha\gamma' \in \alpha\Gamma'$, and since left multiplication by γ permutes the cosets $\alpha_j\Gamma'$, it follows that $\{f|[(\gamma\alpha_j)^{-1}]_0\}$ is merely a rearrangement of $\{f|[\alpha_j^{-1}]_0\}$. Thus, $g|[\gamma^{-1}]_0 = g$, and we conclude that $g \in M_0(\Gamma)$. By Proposition 9(a), g is a constant. Since the term in (3.11) corresponding to the coset $I\Gamma'$ is $f(z) - a$, it follows that for $z = z_0$ the product (3.11) includes a zero factor. Thus, $g = 0$. Then one of the factors in (3.11) must be the zero function (since the meromorphic functions on H form a field). That is, $f(\alpha_j^{-1}z) - a = 0$ for some j and for all $z \in H$. Replacing z by $\alpha_j z$, we have: $f(z) = a$ for all $z \in H$, as claimed. □

As an example of the applications of Proposition 18, we show that for any positive integers N and k (with k even) such that $k(N+1) = 24$, a

$$c_k \underset{\text{def}}{=} \frac{(-1)^{k-1}2k\zeta(k)}{N^k B_k}; \qquad b_0^{\underline{a}} = b_{0,k}^{\underline{a}} \underset{\text{def}}{=} \begin{cases} 0 & \text{if } a_1 \neq 0; \\ \zeta^{a_2}(k) + (-1)^k \zeta^{-a_2}(k) & \text{if } a_1 = 0. \end{cases} \tag{3.18}$$

Then

$$\begin{aligned} G_k^{\underline{a}}(z) &= b_0^{\underline{a}} + \sum_{m_1 \equiv a_1 \bmod N, m_1 \neq 0} \ \sum_{m_2 \equiv a_2 \bmod N} (m_1 z + m_2)^{-k} \\ &= b_0^{\underline{a}} + N^{-k} \sum_{m_1 \equiv a_1 \bmod N, m_1 > 0} \ \sum_{n \in \mathbb{Z}} \left(\frac{m_1 z + a_2}{N} + n \right)^{-k} \\ &\quad + N^{-k}(-1)^k \sum_{m_1 \equiv -a_1 \bmod N, m_1 > 0} \ \sum_{n \in \mathbb{Z}} \left(\frac{m_1 z - a_2}{N} + n \right)^{-k} \\ &= b_0^{\underline{a}} + c_k \Bigg(\sum_{\substack{m_1 \equiv a_1 \bmod N \\ m_1 > 0}} \sum_{j=1}^{\infty} j^{k-1} e^{2\pi i j((m_1 z + a_2)/N)} \\ &\quad + (-1)^k \sum_{\substack{m_1 \equiv -a_1 \bmod N \\ m_1 > 0}} \sum_{j=1}^{\infty} j^{k-1} e^{2\pi i j((m_1 z - a_2)/N)} \Bigg). \end{aligned}$$

In these computations we had to split up the sum into two parts, with m_1 replaced by $-m_1$ for m_1 negative, because in applying (3.17) with z replaced by $(m_1 z \pm a_2)/N$ we need $(m_1 z \pm a_2)/N \in H$, i.e., $m_1 > 0$. Now let

$$\xi \underset{\text{def}}{=} e^{2\pi i/N}, \qquad q_N = e^{2\pi i z/N}. \tag{3.19}$$

Then

$$\begin{aligned} G_k^{\underline{a}}(z) = b_0^{\underline{a}} + c_k \Bigg(&\sum_{\substack{m_1 \equiv a_1 \bmod N \\ m_1 > 0}} \sum_{j=1}^{\infty} j^{k-1} \xi^{j a_2} q_N^{j m_1} \\ &+ (-1)^k \sum_{\substack{m_1 \equiv -a_1 \bmod N \\ m_1 > 0}} \sum_{j=1}^{\infty} j^{k-1} \xi^{-j a_2} q_N^{j m_1} \Bigg). \end{aligned} \tag{3.20}$$

To find the coefficient $b_n^{\underline{a}}$ of q_N^n, it remains to gather together terms with $jm_1 = n$. We shall only do this in the cases when $a_1 = 0$ and $a_2 = 0$. As a result, we have the following proposition.

Proposition 22. *Let $c_k, b_{0,k}^{\underline{a}}, \xi, q_N$ be as in* (3.18)–(3.19). *For $k \geq 3$ let $G_k^{\underline{a} \bmod N}(z)$ be the Eisenstein series* (3.12). *Then the q_N-expansion of $G_k^{\underline{a} \bmod N}$*

$$G_k^{\underline{a}}(z) = b_{0,k}^{\underline{a}} + \sum_{n=1}^{\infty} b_{n,k}^{\underline{a}} q_N^n \tag{3.21}$$

can be computed from (3.20). *If $\underline{a} = (a_1, 0)$, then for $n \geq 1$*

$$b_{n,k}^{\underline{a}} = c_k \left(\sum_{\substack{j \mid n \\ n/j \equiv a_1 \bmod N}} j^{k-1} + (-1)^k \sum_{\substack{j \mid n \\ n/j \equiv -a_1 \bmod N}} j^{k-1} \right). \tag{3.22}$$

If $\underline{a} = (0, a_2)$, *then for* $n \geq 1$

$$b_{n,k}^{\underline{a}} = 0 \quad \text{if} \quad N \nmid n; \qquad b_{Nn,k}^{\underline{a}} = c_k \sum_{j|n} j^{k-1}(\xi^{ja_2} + (-1)^k \xi^{-ja_2}). \tag{3.23}$$

Thus, for $\underline{a} = (0, a_2)$

$$G_k^{(0,a_2)}(z) = b_{0,k}^{(0,a_2)} + c_k \sum_{n=1}^{\infty} \left(\sum_{j|n} j^{k-1}(\xi^{ja_2} + (-1)^k \xi^{-ja_2}) \right) q^n, \qquad q = e^{2\pi i z}. \tag{3.24}$$

Proposition 23. *If* $2\underline{a} \equiv (0, 0) \bmod N$ *and* k *is odd, then* $G_k^{\underline{a} \bmod N} = 0$. *Otherwise,* $G_k^{(0,a_2)}$ *is nonzero at* ∞, *and* $G_k^{(a_1,0)}$ *has a zero of order* $\min(a_1, N - a_1)$, *where we are taking* a_1 *in the range* $0 < a_1 < N$. *That is, for* $\underline{a} = (a_1, 0)$ *the first power of* q_N *which occurs in* (3.21) *with nonzero coefficient is* $q_N^{a_1}$ *or* $q_N^{N-a_1}$.

Proof. The first assertion we already saw as a result of (3.15). We now check that (3.14) is nonzero (unless $N | 2a_2$ and k is odd). If k is even, then we have a sum of positive terms. If k is odd and we take $0 < a_2 < N$, then the sum in (3.14) is equal to

$$\sum_{n=0}^{\infty} \left(\frac{1}{(a_2 + nN)^k} - \frac{1}{(N - a_2 + nN)^k} \right) \begin{cases} > 0 & \text{if } a_2 < N/2; \\ < 0 & \text{if } a_2 > N/2. \end{cases}$$

Finally, we look for the first possible value of n in (3.22) for which either sum in (3.22) is nonzero. That value is $n = \min(a_1, N - a_1)$, where we have the possible value $j = 1$ in one of the two sums. Thus, $b_{n,k}^{(a_1,0)} = \pm c_k$ for $n = \min(a_1, N - a_1)$. This completes the proof. □

As an application, we show that a certain product of $G_3^{\underline{a} \bmod N}$ can be expressed in terms of the η-function. We shall use this result in the next section, where we give Hecke's proof of the transformation formula for $\Theta(z)$.

Recall the Weierstrass $\wp$-function and its derivatives from Chapter I:

$$\wp(z; \omega_1, \omega_2) = \frac{1}{z^2} + \sum_{m,n \in \mathbb{Z}}' \left(\frac{1}{(z + m\omega_1 + n\omega_2)^2} - \frac{1}{(m\omega_1 + n\omega_2)^2} \right);$$

$$\wp'(z; \omega_1, \omega_2) = -2 \sum_{m,n \in \mathbb{Z}} (z + m\omega_1 + n\omega_2)^{-3};$$

$$\wp^{(k-2)}(z; \omega_1, \omega_2) = (-1)^k (k-1)! \sum_{m,n \in \mathbb{Z}} (z + m\omega_1 + n\omega_2)^{-k}, \qquad k \geq 3.$$

If $\underline{a} \neq (0, 0)$, we can express $G_k^{\underline{a} \bmod N}(z)$ in terms of $\wp^{(k-2)}$ as follows:

$$\begin{aligned} G_k^{\underline{a} \bmod N}(z) &= N^{-k} \sum_{m,n \in \mathbb{Z}} \left(\frac{a_1 z + a_2}{N} + mz + n \right)^{-k} \\ &= \frac{(-1)^k}{N^k (k-1)!} \wp^{(k-2)}\left(\frac{a_1 z + a_2}{N}; z, 1 \right). \end{aligned} \tag{3.25}$$

This is the value of $\wp^{(k-2)}$ for the lattice $L_z = \{mz + n\}$ at a point of order

N modulo the lattice, namely $(a_1 z + a_2)/N$. Thus, the Eisenstein series for $\Gamma(N)$ are related to the values of the derivatives of $\wp$ at division points.

Now suppose that $k = 3$. By (3.25), $G_3^{\underline{a}}(z)$ can vanish only if $\wp'((a_1 z + a_2)/N; z, 1) = 0$. But $\wp'$ vanishes only at half-lattice points (see the end of §I.4). If $2\underline{a} \equiv (0, 0) \bmod N$, i.e., if $2(a_1 z + a_2)/N$ is a point in L_z, then $G_3^{\underline{a}}$ is the zero function, by Proposition 23. Otherwise, $\wp'((a_1 z + a_2)/N; z, 1)$ is nonzero. We have proved

Proposition 24. *If* $2\underline{a} \not\equiv 0 \bmod N$, *then* $G_3^{\underline{a} \bmod N}(z) \neq 0$ *for* $z \in H$.

Let p be an odd prime. We now define

$$h(z) = \prod_{a_2=1}^{p-1} G_3^{(0,a_2) \bmod p}(z). \tag{3.26}$$

Proposition 25. $h(z) \in M_{3(p-1)}(\Gamma_0(p))$, *its only zero is a* $(p^2 - 1)/4$*-fold zero at* 0, *and it is a constant multiple of* $(\eta^p(z)/\eta(pz))^6$.

Proof. The first part is the special case of (3.16) when $N = p$, $k = 3$. $h(z)$ is nonzero on H by Proposition 24, and at infinity by Proposition 23. To find its order of zero at 0, we examine the q_p-expansion of

$$h|[S]_{3(p-1)} = \prod_{a_2=1}^{p-1} G_3^{(0,a_2) \bmod p}|[S]_3 = \prod_{a_2=1}^{p-1} G_3^{(a_2,0) \bmod p}$$

by (3.13). According to Proposition 23, the first power of q_p which appears in $G_3^{(a_2,0) \bmod p}$ is $q_p^{\min(a_2, p-a_2)}$. Thus, the order of zero of $h(z)$ at 0 is

$$\sum_{a_2=1}^{p-1} \min(a_2, p - a_2) = 2 \sum_{a_2=1}^{(p-1)/2} a_2 = (p^2 - 1)/4.$$

Now set $\tilde{h}(z) = (\eta^p(z)/\eta(pz))^6$. Then $\tilde{h}(z)^4 = \Delta(z)^p/\Delta(pz)$ is holomorphic and nonzero on H, since $\Delta(z)$ is holomorphic and nonzero on H. Because $\Delta(z) \in S_{12}(\Gamma) \subset S_{12}(\Gamma_0(p))$ and $\Delta(pz) \in S_{12}(\Gamma_0(p))$ by Proposition 17(a), it follows that $\tilde{h}(z)^4$ is a modular function of weight $12(p-1)$ for $\Gamma_0(p)$. Its q-expansion at infinity is

$$q^p \prod (1 - q^n)^{24p}/q^p \prod (1 - q^{np})^{24} = \prod ((1 - q^n)^p/(1 - q^{np}))^{24};$$

hence $\tilde{h}(z)^4$ is holomorphic and nonzero at ∞. At the cusp 0 we have

$$\begin{aligned}
\tilde{h}(z)^4|[S]_{12(p-1)} &= z^{-12(p-1)}\Delta(-1/z)^p/\Delta(-p/z) \\
&= z^{-12(p-1)} z^{12p}\Delta(z)^p/((z/p)^{12}\Delta(z/p)) \\
&= p^{12}\Delta(z)^p/\Delta(z/p) = p^{12} q^p \prod (1 - q^n)^{24p}/q_p \prod (1 - q_p^n)^{24},
\end{aligned}$$

which has leading term $p^{12} q_p^{p^2-1}$. Thus, both h^4 and $\tilde{h}^4$ are elements of $M_{12(p-1)}(\Gamma_0(p))$ with no zero except for a $(p^2 - 1)$-order zero at 0. Hence their ratio is a constant by Proposition 18. But $(h/\tilde{h})^4 = \text{const}$ implies that $h/\tilde{h} = \text{const}$. This concludes the proof of Proposition 25. □

Notice that, by (3.15), we have some duplication in the definition (3.26) of $h(z)$. That is, if we define

$$f(z) = \prod_{a_2=1}^{(p-1)/2} G_3^{(0,a_2) \bmod p}(z), \tag{3.27}$$

we find, by (3.15), that

$$h(z) = (-1)^{(p-1)/2} f(z)^2. \tag{3.28}$$

Proposition 25 then tells us that the square of the ratio of $f(z)$ to $(\eta^p(z)/\eta(pz))^3$ is a constant. Hence, the ratio of those two functions must itself be a constant, and we have proved

Proposition 26. *The function* $f(z)$ *defined in* (3.27) *is a constant multiple of* $(\eta^p(z)/\eta(pz))^3$.

Because each of the $G_3^{(0,a_2)}$ in (3.27) is in $M_3(\Gamma_1(p))$ by Proposition 21, it follows that $f \in M_{3(p-1)/2}(\Gamma_1(p))$. However, unlike $h(z)$, $f(z)$ is not, strictly speaking, a modular form for the larger group $\Gamma_0(p)$.

Proposition 27. *Let* $f(z)$ *be defined by* (3.27), *and let* $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(p)$. *Then* $f|[\gamma]_{3(p-1)/2} = \left(\frac{d}{p}\right) f$, *where* $\left(\frac{d}{p}\right)$ *is the Legendre symbol* (*which is* ± 1, *depending on whether or not d is a square modulo p*).

PROOF. Since $(0, a_2)\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv (0, da_2) \bmod p$, it follows by (3.13) that

$$f|[\gamma]_{3(p-1)/2} = \prod_{a_2=1}^{(p-1)/2} G_3^{(0,da_2) \bmod p}.$$

But by (3.15), the terms in this product are a rearrangement of (3.27), except that a minus sign is introduced every time the least positive residue of da_2 modulo p falls in the range $(p+1)/2, (p+3)/2, \ldots, p-1$. Let n_d be the number of times this occurs. Thus, $f|[\gamma]_{3(p-1)/2} = (-1)^{n_d} f$. According to Gauss's lemma, which is an easily proved fact from elementary number theory (see, for example, p. 74 of [Hardy and Wright 1960]), we have $(-1)^{n_d} = \left(\frac{d}{p}\right)$. This proves the proposition. □

The transformation formula in Proposition 27 is an example of a general relationship between modular forms for $\Gamma_1(N)$ and "twisted-modular" forms for $\Gamma_0(N)$, called "modular forms with character". We now discuss this relationship. We start with some very general observations.

Suppose that Γ'' is a subgroup of Γ', and $f(z)$ is a modular form of weight k for the smaller group Γ'' but not necessarily for the bigger one. Then for $\gamma \in \Gamma'$ we can at least say that $f|[\gamma^{-1}]_k$ only depends on the coset of γ modulo Γ''. That is, if $\gamma'' \in \Gamma''$, then $f|[(\gamma\gamma'')^{-1}]_k = f|[\gamma^{-1}]_k$.

Now suppose that the subgroup Γ'' is normal in Γ'. To every coset $\gamma\Gamma''$ associate the linear map $f \mapsto f|[\gamma^{-1}]_k$ which takes an element in $M_k(\Gamma'')$ to $M_k(\Gamma'')$ by Proposition 17 (with Γ'' and γ^{-1} in place of Γ' and α; note that $\gamma\Gamma''\gamma^{-1} \cap \Gamma = \Gamma''$, since $\gamma \in \Gamma'$ and Γ'' is normal in Γ'). This gives us a group

homomorphism ρ from Γ'/Γ'' to the linear automorphisms of the vector space $M_k(\Gamma'')$, because for $\gamma_1, \gamma_2 \in \Gamma'$

$$\rho(\gamma_1\gamma_2): f \mapsto f|[(\gamma_1\gamma_2)^{-1}]_k = (f|[\gamma_2^{-1}]_k)|[\gamma_1^{-1}]_k = \rho(\gamma_1)(\rho(\gamma_2)f).$$

In other words, we have what is called a "representation" of the group Γ'/Γ'' in the vector space $M_k(\Gamma'')$.

If $\chi: \Gamma'/\Gamma'' \to \mathbb{C}^*$ is a character of the quotient group, then define $M_k(\Gamma', \chi)$ to be the subspace of $M_k(\Gamma'')$ consisting of modular forms on which the representation ρ acts by scalar multiplication by χ, i.e.,

$$M_k(\Gamma', \chi) \underset{\text{def}}{=} \{f \in M_k(\Gamma'') | \rho(\gamma)f = \chi(\gamma)f \text{ for all } \gamma \in \Gamma'\}.$$

If Γ'/Γ'' happens to be abelian, then according to a basic fact about representations of finite abelian groups, $M_k(\Gamma'')$ decomposes into a direct sum of $M_k(\Gamma', \chi)$ over the various characters χ of Γ'/Γ''. We shall soon recall the simple proof of this fact in the special case that will interest us.

We apply these observations to the case $\Gamma'' = \Gamma_1(N)$, $\Gamma' = \Gamma_0(N)$. Since $\Gamma_1(N)$ is the kernel of the surjective homomorphism from $\Gamma_0(N)$ to $(\mathbb{Z}/N\mathbb{Z})^*$ that takes $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ to d, it follows that $\Gamma_1(N)$ is a normal subgroup of $\Gamma_0(N)$ with abelian quotient group isomorphic to $(\mathbb{Z}/N\mathbb{Z})^*$ (see Problems 1–2 of §III.1). Let χ be any Dirichlet character modulo N, i.e., any character of $(\mathbb{Z}/N\mathbb{Z})^*$. In this context the subspace $M_k(\Gamma_0(N), \chi) \subset M_k(\Gamma_1(N))$ is usually abbreviated $M_k(N, \chi)$. That is,

$$M_k(N, \chi) \underset{\text{def}}{=} \left\{ f \in M_k(\Gamma_1(N)) | f|[\gamma]_k = \chi(d)f \quad \text{for } \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N) \right\}. \tag{3.29}$$

In particular, if $\chi = 1$ is the trivial character, then $M_k(N, 1) = M_k(\Gamma_0(N))$.

Proposition 28. $M_k(\Gamma_1(N)) = \oplus M_k(N, \chi)$, *where the sum is over all Dirichlet characters modulo N.*

Proof. As mentioned before, this proposition is actually a special case of the basic fact from representation theory that any representation of a finite abelian group decomposes into a direct sum of characters. However, we shall give an explicit proof anyway.

First, any function f that satisfies $f|[\gamma]_k = \chi(d)f$ for two distinct characters χ must clearly be zero; hence, it suffices to show that any $f \in M_k(\Gamma_1(N))$ can be written as a sum of functions $f_\chi \in M_k(N, \chi)$. Let

$$f_\chi = \frac{1}{\phi(N)} \sum_{d \in (\mathbb{Z}/N\mathbb{Z})^*} \bar{\chi}(d) f|[\gamma_d]_k,$$

where γ_d is any element of $\Gamma_0(N)$ with lower-right entry congruent to d mod N. We check that for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d' \end{smallmatrix}\right) \in \Gamma_0(N)$

$$f_\chi|[\gamma]_k = \frac{1}{\phi(N)} \sum_{d \in (\mathbb{Z}/N\mathbb{Z})^*} \bar{\chi}(d) f|[\gamma_{dd'}]_k,$$

which, if we replace dd' by d as the variable of summation, is easily seen

to equal $\chi(d')f_\chi$, i.e., $f_\chi \in M_k(N, \chi)$. Finally, we sum the f_χ over all characters χ modulo N, reverse the order of summation over d and χ, and obtain:

$$\sum f_\chi = \sum_{d \in (\mathbb{Z}/N\mathbb{Z})^*} \frac{1}{\phi(N)} \sum_\chi \bar{\chi}(d) \; f|[\gamma_d]_k,$$

which is equal to f, because the inner sum is 1 if $d = 1$ and 0 otherwise. Thus, f can be written as a sum of functions in $M_k(N, \chi)$, as claimed. □

Notice that $M_k(N, \chi) = 0$ if χ has a different parity from k, i.e., if $\chi(-1) \neq (-1)^k$. This follows by taking $\gamma = -I$ in the definition (3.29) and recalling that $f|[-I]_k = (-1)^k f$.

For example, as an immediate corollary of Proposition 28 and the preceding remark, we have

Proposition 29.

$$M_k(\Gamma_1(4)) = \begin{cases} M_k(4, 1), & k \text{ even;} \\ M_k(4, \chi), & k \text{ odd,} \end{cases}$$

where 1 *denotes the trivial character and* χ *the unique nontrivial character modulo* 4.

Notice that the relationship in (3.29) is multiplicative in γ; that is, if it holds for γ_1 and γ_2, then it holds for their product. Thus, as in the case of modular forms without character, to show that $f(z)$ is in $M_k(N, \chi)$ it suffices to check the transformation rule on a set of elements that generate $\Gamma_0(N)$.

As another example, we look at $\Theta^2(z) = (\Sigma_{n \in \mathbb{Z}} q^{n^2})^2$, whose n-th q-expansion coefficient is the number of ways n can be written as a sum of two squares.

Proposition 30. $\Theta^2 \in M_1(\Gamma_1(4)) = M_1(4, \chi)$, *where* $\chi(d) = (-1)^{(d-1)/2}$.

Proof. It suffices to verify the transformation rule for $-I$, T, and $ST^4S = \begin{pmatrix} -1 & 0 \\ 4 & -1 \end{pmatrix}$, which generate $\Gamma_0(4)$ (see Problem 13 of §III.1). This is immediate for T, since Θ^2 has period 1. Next, the relation $f|[-I]_1 = -f = \chi(-1)f$ holds for any f, by definition. So it remains to treat the case ST^4S. Let

$$\alpha_N \underset{\text{def}}{=} \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}, \quad \text{so that} \quad \alpha_N^{-1} = -\frac{1}{N}\alpha_N,$$
$$\text{and} \quad \alpha_N \begin{pmatrix} a & b \\ c & d \end{pmatrix} \alpha_N^{-1} = \begin{pmatrix} d & -c/N \\ -Nb & a \end{pmatrix}. \tag{3.30}$$

We write $ST^4S = -\alpha_4 T \alpha_4^{-1} = \frac{1}{4}\alpha_4 T \alpha_4$, and use the relationship $\Theta^2|[\alpha_4]_1 = -i\Theta^2$ (see (3.5)) to obtain

$$\Theta^2|[ST^4S]_1 = \Theta^2|[\alpha_4 T \alpha_4]_1 = -i\Theta^2|[T\alpha_4]_1 = -i\Theta^2|[\alpha_4]_1 = -\Theta^2.$$

(Recall that the scalar matrix $\frac{1}{4}I$ acts trivially on all functions, i.e., $[1/4]_1 =$ identity.)

To finish the proof of the proposition, we must show the cusp condition, i.e., that $\Theta^2|[\gamma_0]_1$ is finite at infinity for all $\gamma_0 \in \Gamma$. But the square of $\Theta^2|[\gamma_0]_1$ is $\Theta^4|[\gamma_0]_2$, and it will be shown in Problem 11 below that $\Theta^4 \in M_2(\Gamma_0(4))$; in particular, this means that $\Theta^4|[\gamma_0]_2$, and hence also $\Theta^2|[\gamma_0]_1$, is finite at infinity. This completes the proof. □

The spaces $M_k(N, \chi)$ include many of the most important examples of modular forms, and will be our basic object of study in several of the sections that follow. We also introduce the notation $S_k(N, \chi)$ to denote the subspace of cusp forms: $S_k(N, \chi) \underset{\text{def}}{=} M_k(N, \chi) \cap S_k(\Gamma_1(N))$.

The Mellin transform of a modular form. Suppose that $f(z) = \Sigma\, a_n q_N^n$ (where $q_N = e^{2\pi i z/N}$) is a modular form of weight k for a congruence subgroup Γ' of level N. Further suppose that $|a_n| = O(n^c)$ for some constant $c \in \mathbb{R}$, i.e., that a_n/n^c is bounded as $n \to \infty$. It is not hard to see that the q_N-expansion coefficients for the Eisenstein series $G_k^{\underline{a} \bmod N}$ have this property with $c = k - 1 + \varepsilon$ for any $\varepsilon > 0$. For example, in the case $\Gamma' = \Gamma$, the coefficients are a constant multiple of $\sigma_{k-1}(n)$, and it is not hard to show that $\sigma_{k-1}(n)/n^{k-1+\varepsilon} \to 0$ as $n \to \infty$. We shall later show that, if f is a cusp form, we can take $c = k/2 + \varepsilon$. It has been shown (as a consequence of Deligne's proof of the Weil conjectures) that one can actually do better, and take $c = (k-1)/2 + \varepsilon$.

In Chapter II we saw that the Mellin transform of $\theta(t) = \Sigma\, e^{-\pi t n^2}$ and certain generalizations are useful in investigating some important Dirichlet series, such as the Riemann zeta-function, Dirichlet L-functions, and the Hasse–Weil L-function of the elliptic curves E_n: $y^2 = x^3 - n^2 x$. We now look at the Mellin transform for modular forms.

Because we use a variable z in the upper half-plane rather than t (e.g., $t = -2iz$), we define the Mellin transform by integrating along the positive imaginary axis rather than the positive real axis.

The most important case is $\Gamma' = \Gamma_1(N)$. For now we shall also assume that $f(\infty) = 0$. Thus, let $f(z) = \Sigma_{n=1}^{\infty} a_n q^n \in M_k(\Gamma_1(N))$. (Recall that since $T \in \Gamma_1(N)$, we have an expansion in powers of $q = e^{2\pi i z}$ rather than q_N.) We set

$$g(s) \underset{\text{def}}{=} \int_0^{i\infty} f(z) z^{s-1} dz. \tag{3.31}$$

We now show that if $f(z) = \Sigma_{n=1}^{\infty} a_n q^n$ with $|a_n| = O(n^c)$, then the integral $g(s)$ defined in (3.31) converges for $\operatorname{Re} s > c + 1$:

$$\begin{aligned}\int_0^{i\infty} f(z) z^{s-1} dz &= \sum_{n=1}^{\infty} a_n \int_0^{i\infty} z^s e^{2\pi i n z} \frac{dz}{z} \\ &= \sum_{n=1}^{\infty} a_n \left(-\frac{1}{2\pi i n}\right)^s \int_0^{\infty} t^s e^{-t} \frac{dt}{t} \qquad \text{(where } t = -2\pi i n z\text{)} \\ &= (-2\pi i)^{-s} \Gamma(s) \sum_{n=1}^{\infty} a_n n^{-s} \qquad \text{(see (4.6) of Ch. II)}\end{aligned} \tag{3.32}$$

(where the use of Γ in the gamma-function $\Gamma(s)$ has no relation to its use in the notation for congruence subgroups; but in practice the use of the same letter Γ should not cause any confusion). Since $|a_n n^{-s}| = O(n^{c-\operatorname{Re} s})$, this last sum is absolutely convergent (and the interchanging of the order of integration and summation was justified).

If $f(z) = \Sigma_{n=0}^{\infty} a_n q^n \in M_k(\Gamma_1(N))$ has $a_0 \neq 0$, we replace $f(z)$ by $f(z) - a_0$ in (3.31). In either case, we then obtain $g(s) = (-2\pi i)^{-s}\Gamma(s)L_f(s)$, where

$$L_f(s) \underset{\text{def}}{=} \sum_{n=1}^{\infty} a_n n^{-s} \quad \text{for} \quad \operatorname{Re} s > c+1, \quad \text{if} \quad f(z) = \sum_{n=0}^{\infty} a_n q^n \quad \text{with} \quad |a_n| = O(n^c). \tag{3.33}$$

In addition to their invariance under $[\gamma]_k$ for $\gamma \in \Gamma_1(N)$, many modular forms also transform nicely under $[\alpha_N]_k$, where $\alpha_N = \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right)$ as in (3.30). It will be shown in the exercises that, for example, any function in $M_k(N, \chi)$ for χ a real character (i.e., its values are ± 1) can be written as a sum of two functions satisfying

$$f|[\alpha_N]_k = Ci^{-k}f, \qquad C = 1 \text{ or } -1, \tag{3.34}$$

where one of the functions satisfies (3.34) with $C = 1$ and the other with $C = -1$. An example we already know of a function satisfying (3.34) is Θ^2: the relation (3.5) is a special case of (3.34) with $k = 1$, $C = 1$, $N = 4$.

We now show that if (3.34) holds, then we have a functional equation for the corresponding Mellin transform $g(s)$ which relates $g(s)$ to $g(k-s)$. For simplicity, we shall again suppose that $f(z) = \Sigma a_n q^n$ with $a_0 = 0$. We can write (3.34) explicitly as follows, by the definition of $[\alpha_N]_k$:

$$f(-1/Nz) = CN^{-k/2}(-iNz)^k f(z). \tag{3.35}$$

In (3.31), we break up the integral into the part from 0 to $i/\sqrt{N}$ and the part from $i/\sqrt{N}$ to $i\infty$. We choose $i/\sqrt{N}$ because it is the fixed point in H of $\alpha_N: z \mapsto -1/Nz$. We have

$$\begin{aligned} g(s) &= \int_{i/\sqrt{N}}^{i\infty} f(z)z^s \frac{dz}{z} - \int_{i/\sqrt{N}}^{i\infty} f(-1/Nz)(-1/Nz)^s \frac{d(-1/Nz)}{-1/Nz} \\ &= \int_{i/\sqrt{N}}^{i\infty} (f(z)z^s + f(-1/Nz)(-1/Nz)^s)\frac{dz}{z} \\ &= \int_{i/\sqrt{N}}^{i\infty} (f(z)z^s + i^k CN^{-k/2} f(z)(-1/Nz)^{s-k})\frac{dz}{z}, \end{aligned}$$

because of (3.35).

In the first place, this integral converges to an entire function of s, because $f(z)$ decreases exponentially as $z \to i\infty$. That is, because the lower limit of integration has been moved away from zero, we no longer have to worry about the behavior of the integrand near 0. (Compare with the proof of Proposition 13 in Chapter II, where we used a similar technique to find a

rapidly convergent series for the critical value of the Hasse–Weil L-function; see the remark following equation (6.7) in §II.6.)

Moreover, if we replace s by $k - s$ in the last integral and factor out $i^k C N^{-k/2}(-N)^s$, we obtain:

$$g(k-s) = i^k C N^{-k/2}(-N)^s \int_{i/\sqrt{N}}^{i\infty} (i^{-k} C N^{k/2}(-N)^{-s} f(z) z^{k-s} + f(z) z^s) \frac{dz}{z}$$

$$= i^k C N^{-k/2}(-N)^s \int_{i/\sqrt{N}}^{i\infty} (i^k C N^{-k/2} f(z)(-1/Nz)^{s-k} + f(z) z^s) \frac{dz}{z}$$

$$= i^k C N^{-k/2}(-N)^s g(s),$$

because the last integral is the same as our earlier integral for $g(s)$. This equality can be written in the form

$$(-i\sqrt{N})^s g(s) = C(-i\sqrt{N})^{k-s} g(k-s).$$

Thus, by (3.32)–(3.33), if we define $\Lambda(s)$ for $\operatorname{Re} s > c + 1$ by

$$\Lambda(s) = (-i\sqrt{N})^s g(s) = (\sqrt{N}/2\pi)^s \Gamma(s) L_f(s), \tag{3.36}$$

we have shown that $\Lambda(s)$ extends to an entire function of s, and satisfies the functional equation

$$\Lambda(s) = C\Lambda(k-s). \tag{3.37}$$

As an example of this result, we can take $f(z) = \Delta(z) \in S_{12}(\Gamma)$, which satisfies (3.35) with $N = 1$, $k = 12$, $C = 1$. Then $\Delta(z) = \Sigma_{n=1}^{\infty} \tau(n) q^n$, $L_\Delta(s) = \Sigma_{n=1}^{\infty} \tau(n) n^{-s}$, and $\Lambda(s) = (2\pi)^{-s}\Gamma(s) L_\Delta(s)$ satisfies the relation: $\Lambda(s) = \Lambda(12 - s)$.

The derivation of (3.37) from (3.34) indicates a close connection between Dirichlet series with a functional equation and modular forms. We came across Dirichlet series with a functional equation in a very different context in Chapter II. Namely, the Hasse–Weil L-function of the elliptic curve E_n: $y^2 = x^3 - n^2 x$ satisfies (3.36)–(3.37) with $k = 2$, $N = 32n^2$ for n odd and $16n^2$ for n even, $C = (\frac{-2}{n})$ for n odd and $(\frac{-2}{n/2})$ for n even (see (5.10)–(5.12) in Ch. II). We also saw that the Hasse–Weil L-function of the elliptic curve $y^2 = x^3 + 16$ satisfies (3.36)–(3.37) with $k = 2$, $N = 27$, $C = 1$ (see Problem 8(d) of §II.5).

So the question naturally arises: Can one go the other way? Does every Dirichlet series with the right type of functional equation come from some modular form, i.e., is it of the form $L_f(s)$ for some modular form f? In particular, can the Hasse–Weil L-functions we studied in Chapter II be obtained by taking the Mellin transform of a suitable modular form of weight 2? That is, if we write $L(E_n, s)$ in the form $\Sigma_{m=1}^{\infty} b_m m^{-s}$ (see (5.3) in Ch. II), is $\Sigma_{m=1}^{\infty} b_m q^m$ the q-expansion of a weight two modular form?

Hecke [1936] and Weil [1967] showed that the answer to these questions is basically yes, but with some qualifications. We shall not give the details,

which are available in [Ogg 1969], but shall only outline the situation and state Weil's fundamental theorem on the subject.

Suppose that $L(s) = \Sigma a_n n^{-s}$ satisfies (3.36)–(3.37) (and a suitable hypothesis about convergence). Using the "inverse Mellin transform", one can reverse the steps that led to (3.36)–(3.37), and find that $f(z) = \Sigma a_n q^n$ satisfies (3.34). For now, let us suppose that $N = \lambda^2$ is a perfect square, and that $C = i^k$ (k even). Then, if $f(z)$ satisfies (3.35), it follows that $f_1(z) \underset{\text{def}}{=} f(z/\lambda) = \Sigma a_n q_\lambda^n$ satisfies: $f_1(-1/z) = z^k f_1(z)$. Thus, f_1 is invariant under $[S]_k$ and $[T^\lambda]_k$, and hence is invariant under the group generated by S and T^λ. Hecke denoted that group $\mathfrak{G}(\lambda)$. We have encountered the group $\mathfrak{G}(2)$ before.

In this way one can show, for example, that $L(E_{2n_0}, s)$ corresponds to a modular form (actually, a cusp form) of weight 2 for $\mathfrak{G}(8n_0)$.

Unfortunately, however, Hecke's groups $\mathfrak{G}(\lambda)$ turn out not to be large enough to work with satisfactorily. In general, they are not congruence subgroups. ($\mathfrak{G}(2) \supset \Gamma(2)$ is an exception.)

But one can do much better. Weil showed, roughly speaking, that if one has functional equations analogous to (3.36)–(3.37) for enough "twists" $\Sigma \chi(n) a_n n^{-s}$ of the Dirichlet series $\Sigma a_n n^{-s}$, then the corresponding q-expansion is in $M_k(\Gamma_0(N))$. We now give a more precise statement of Weil's theorem.

Let χ_0 be a fixed Dirichlet character modulo N (χ_0 is allowed to be the trivial character). Let χ be a variable Dirichlet character of conductor m, where m is either an odd prime not dividing N, or else 4 (we allow $m = 4$ only if N is odd). By a "large" set of values of m we shall mean that the set contains at least one m in any given arithmetic progression $\{u + jv\}_{j \in \mathbb{Z}}$, where u and v are relatively prime. According to Dirichlet's theorem, any such arithmetic progression contains a prime; thus, a "large" set of primes is one which satisfies (this weak form of) Dirichlet's theorem. By a "large" set of characters χ we shall mean the set of all nontrivial χ modulo m for a "large" set of m.

Let $C = \pm 1$, and for any χ of conductor m set

$$C_\chi = C\chi_0(m)\chi(-N)g(\chi)/g(\bar{\chi}), \tag{3.38}$$

where $g(\chi) = \Sigma_{j=1}^m \chi(j) e^{2\pi i j/N}$ is the Gauss sum. Given a q-expansion $f(z) = \Sigma_{n=0}^\infty a_n q^n$, $q = e^{2\pi i z}$, for which $|a_n| = O(n^c)$, we define $L_f(s)$ by (3.33) and $\Lambda(s)$ by (3.36), and we further define

$$L_f(\chi, s) = \sum_{n=1}^{\infty} \chi(n) a_n n^{-s}; \qquad \Lambda(\chi, s) = (m\sqrt{N}/2\pi)^s \Gamma(s) L_f(\chi, s). \tag{3.39}$$

Weil's Theorem. *Suppose that* $f(z) = \Sigma_{n=0}^\infty a_n q^n$, $q = e^{2\pi i z}$, *has the property that* $|a_n| = O(n^c)$, $c \in \mathbb{R}$. *Suppose that for* $C = 1$ *or* -1 *the function* $\Lambda(s)$ *defined by* (3.36) *has the property that* $\Lambda(s) + a_0(1/s + C/(k-s))$ *extends*

to an entire function which is bounded in any vertical strip of the complex plane, and satisfies the functional equation $\Lambda(s) = C\Lambda(k - s)$. *Further suppose that for a "large" set of characters* χ *of conductor* m (*in the sense explained above*), *the function* $\Lambda(\chi, s)$ *defined by* (3.39) *extends to an entire function which is bounded in any vertical strip, and satisfies the functional equation* $\Lambda(\chi, s) = C_\chi \Lambda(\bar{\chi}, k - s)$, *with* C_χ *defined in* (3.38).

Then $f \in M_k(N, \chi_0)$, *and* f *satisfies* (3.34). *If, in addition,* $L_f(s)$ *converges absolutely for* Re $s > k - \varepsilon$ *for some* $\varepsilon > 0$, *then* f *is a cusp form.*

One can show that the Hasse–Weil L-functions of the elliptic curves in Chapter II satisfy the hypotheses of Weil's theorem (with $\chi_0 = 1$). The same techniques as in the proof of the theorem in §II.5 can be used to show this. However, one must consider the Hecke L-series obtained in (5.6) of Ch. II by replacing $\tilde{\chi}_n(I)$ by the character $\tilde{\chi}_n(I)\chi(\mathbb{N}I)$ with χ any Dirichlet character modulo m as in Weil's theorem. For example, if we do this for $L(E_1, s)$, where E_1 is the elliptic curve $y^2 = x^3 - x$, we can conclude by Weil's theorem that

$$f_{E_1}(z) = q - 2q^5 - 3q^9 + 6q^{13} + 2q^{17} + \sum_{m \geq 25} b_m q^m \tag{3.40}$$

(see (5.4) of Ch. II) is a cusp form of weight two for $\Gamma_0(32)$.

If we form the q-expansion corresponding to the L-series of E_n: $y^2 = x^3 - n^2 x$, namely, $f_{E_n}(z) = \Sigma\, \chi_n(m) b_m q^m$, it turns out that $f_{E_n} \in M_2(\Gamma_0(32n^2))$ for n odd and $f_{E_n} \in M_2(\Gamma_0(16n^2))$ for n even. Note that when $n \equiv 1 \bmod 4$, so that χ_n is a character of conductor n, this is an immediate consequence of the fact that $f_{E_1} \in M_2(\Gamma_0(32))$, by Proposition 17(b).

More generally, it can be shown that the Hasse–Weil L-function for any elliptic curve with complex multiplication satisfies the hypotheses of Weil's theorem with $k = 2$, and so corresponds to a weight two modular form (actually, a cusp form) for $\Gamma_0(N)$. (N is the so-called "conductor" of the elliptic curve.)

Many elliptic curves without complex multiplication are also known to have this property. In fact, it was conjectured (by Taniyama and Weil) that every elliptic curve defined over the rational numbers has L-function which satisfies Weil's theorem for some N. Geometrically, the cusp forms of weight two can be regarded as holomorphic differential forms on the Riemann surface $\Gamma_0(N)\backslash H$ (i.e., the fundamental domain with $\Gamma_0(N)$-equivalent boundary sides identified and the cusps included). The Taniyama–Weil conjecture then can be shown to take the form: every elliptic curve over $\mathbb{Q}$ can be obtained as a quotient of the Jacobian of some such Riemann surface.

For more information about the correspondence between modular forms and Dirichlet series, see [Hecke 1981], [Weil 1967], [Ogg 1969], and [Shimura 1971].

Finally, we mention a dramatic and surprising result of G. Frey, J.-P. Serre, and K. Ribet: *The Taniyama–Weil conjecture implies Fermat's Last Theorem.* For an account, see [Oesterlé 1988].

PROBLEMS

1. Let $\alpha \in GL_2^+(\mathbb{Q})$, and let $g(z) = f(\alpha z)$. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$. Notice that $f(\alpha z)|[\gamma]_k$ was defined to be $(c\alpha z + d)^{-k} f(\gamma\alpha z)$, which is *not* the same as $g(z)|[\gamma]_k = (cz + d)^{-k} f(\alpha\gamma z)$. Show that if $\alpha = \left(\begin{smallmatrix} n & 0 \\ 0 & 1 \end{smallmatrix}\right)$, i.e., if $\alpha z = nz$, then $g(z)|[\gamma]_k = f(\alpha z)|[\alpha\gamma\alpha^{-1}]_k = f(nz)|[\left(\begin{smallmatrix} a & nb \\ c/n & d \end{smallmatrix}\right)]_k$.

2. Let Γ' be a congruence subgroup of Γ of level N, and denote $\Gamma'_s = \{\gamma \in \Gamma' | \gamma s = s\}$ for $s \in \mathbb{Q} \cup \{\infty\}$. Let $s = \alpha^{-1}\infty$, $\alpha \in \Gamma$.
   (a) Prove that $\alpha\Gamma'_s\alpha^{-1} = (\alpha\Gamma'\alpha^{-1})_\infty$.
   (b) Show that there exists a unique positive integer h (called the "ramification index" of Γ' at s) such that

   (i) in the case $-I \in \Gamma'$
   $$\Gamma'_s = \pm\alpha^{-1}\{T^{hm}\}_{n \in \mathbb{Z}}\alpha;$$
   (ii) in the case $-I \notin \Gamma'$ either
   $$\Gamma'_s = \alpha^{-1}\{T^{hn}\}_{n \in \mathbb{Z}}\alpha; \text{ or} \qquad \text{(IIa)}$$
   $$\Gamma'_s = \alpha^{-1}\{(-T^h)^n\}_{n \in \mathbb{Z}}\alpha. \qquad \text{(IIb)}$$

   Show that h is a divisor of N.
   (c) Show that the integer h and the type (I, IIa, or IIb) of s does not depend on the choice of $\alpha \in \Gamma$ with $s = \alpha^{-1}\infty$; and they only depend on the Γ'-equivalence class of s.
   (d) Show that if $\alpha^{-1}\infty$ is of type I or IIa and $f \in M_k(\Gamma')$, then $f|[\alpha^{-1}]_k$ has a Fourier expansion in powers of q_h. A cusp of Γ' is called "regular" if it is of type I or IIa; it is called "irregular" if it is of type IIb.
   (e) Show that if $\alpha^{-1}\infty$ is an irregular cusp, and $f \in M_k(\Gamma')$, then $f|[\alpha^{-1}]_k$ has a Fourier expansion in powers of q_{2h} in which only odd powers appear if k is odd and only even powers appear if k is even. If k is odd, note that this means that to show that $f \in M_k(\Gamma')$ is a cusp form one need only check the q-expansions at the regular cusps.

3. Let h be any positive integer, and suppose $2h|N$, $N \geq 4$. Let Γ' be the following level N congruence subgroup: $\Gamma' = \{\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} -1 & -h \\ 0 & -1 \end{smallmatrix}\right)^j \bmod N \text{ for some } j\}$. Show that ∞ is a cusp of type IIb.

4. (a) Show that $\Gamma_1(N)$ has the same cusps as $\Gamma_0(N)$ for $N = 3, 4$.
   (b) Note that $-I \notin \Gamma_1(N)$ for $N > 2$. Which of the cusps of $\Gamma_1(3)$ and $\Gamma_1(4)$, if any, are irregular?

5. Find the ramification indices of Γ' at all of its cusps when:
   (a) $\Gamma' = \Gamma_0(p)$ (p a prime);
   (b) $\Gamma' = \Gamma_0(p^2)$;
   (c) $\Gamma' = \Gamma(2)$.

6. Prove that if $\Gamma' \subset \Gamma$ is a normal subgroup, then all cusps have the same ramification index, namely $[\Gamma_\infty : \pm\Gamma'_\infty]$.

7. (a) Show that any weight zero modular function for $\Gamma' \subset \Gamma$ satisfies a polynomial of degree $[\Gamma : \Gamma']$ over the field $\mathbb{C}(j)$ of weight zero modular functions for Γ.
(b) Show that, if Γ' is a *normal* subgroup, and if $f(z)$ is Γ'-invariant, then so is $f(\alpha z)$ for any $\alpha \in \Gamma$. Then show that the field of weight zero modular functions for Γ' is a galois field extension of $\mathbb{C}(j)$ whose galois group is a quotient of Γ/Γ'. In practice (e.g., if Γ' is a congruence subgroup), it can be shown that the galois group is equal to Γ/Γ'.

8. Prove the following identities, which will be useful in the problems that follow and in the next section, by manipulation of power series and products in $q = e^{2\pi i z}$:
(a) $\Theta(z) + \Theta(z + \frac{1}{2}) = 2\Theta(4z)$;
(b) $E_2(z + \frac{1}{2}) - E_2(z) = 48 \sum_{\text{odd } n>0} \sigma_1(n) q^n$;
(c) $E_k(z) - (1 + p^{k-1}) E_k(pz) + p^{k-1} E_k(p^2 z) = -\frac{2k}{B_k} \sum_{p \nmid n} \sigma_{k-1}(n) q^n$ $(k \geq 2, p \text{ prime})$;
(d) $E_2(z) - 3E_2(2z) + 2E_2(4z) = \frac{1}{2}(E_2(z) - E_2(z + \frac{1}{2}))$;
(e) $\eta(z + \frac{1}{2}) = e^{2\pi i/48} \eta^3(2z)/\eta(z)\eta(4z)$.

9. Prove that if k is even and $f(z)$ has period one and satisfies $f(-1/4z) = (-4z^2)^{k/2} f(z)$, then $f|[\gamma]_k = f$ for all $\gamma \in \Gamma_0(4)$.

10. (a) Prove that $\eta^8(4z)/\eta^4(2z) \in M_2(\Gamma_0(4))$, and find its value at each cusp.
(b) For $a \in \mathbb{Z}$ prove that $E_2(ST^{-a}Sz) = (az + 1)^2 E_2(z) - \frac{6ai}{\pi}(az + 1)$.
(c) Let $F(z) \underset{\text{def}}{=} -\frac{1}{24}(E_2(z) - 3E_2(2z) + 2E_2(4z)) = \sum_{\text{odd } n>0} \sigma_1(n) q^n$ by Problem 8(c). Prove that $F(z) \in M_2(\Gamma_0(4))$, and find its value at each cusp.
(d) Prove that $F(z) = \eta^8(4z)/\eta^4(2z)$. Then derive the identity

$$q \prod_{n=1}^{\infty} (1 - q^{4n})^4 (1 + q^{2n})^4 = \sum_{\text{odd } n>0} \sigma_1(n) q^n.$$

(e) Give a different proof that $-24F(z) = \frac{1}{2}(E_2(z) - E_2(z + 1/2))$ is in $M_2(\Gamma_0(4))$ by proving that, more generally, $E_2(z) - \frac{1}{N}\sum_{j=0}^{N-1} E_2(z + j/N)$ is in $M_2(\Gamma_0(N^2))$.

11. (a) Prove that $\Theta(z)^4 \in M_2(\Gamma_0(4))$, and find its value at each cusp.
(b) Show that $\Theta(z)^4$ and $F(z)$ (see preceding problem) are linearly independent.
(c) Prove that $\eta^{20}(2z)/\eta^8(z)\eta^8(4z) \in M_2(\Gamma_0(4))$, and find its value at each cusp.
(d) Prove that $\Theta(z) = \eta^5(2z)/\eta^2(z)\eta^2(4z)$.
(e) Prove that $\Theta(z) = e^{-2\pi i/24} \eta^2(z + \frac{1}{2})/\eta(2z)$.

12. Let $N = 7$ or 23, and let $k = 24/(N + 1)$. Let χ be the Legendre symbol $\chi(n) = (\frac{n}{N})$. Prove that any nonzero element of $S_k(N, \chi)$ must be a constant multiple of $(\eta(z)\eta(Nz))^k$.

13. Using Propositions 25–27, prove that $(\eta(z)\eta(3z))^6 \in S_6(\Gamma_0(3))$ and $(\eta(z)\eta(7z))^3 \in S_3(7, \chi)$ where $\chi(n) = (\frac{n}{7})$.

14. Let $\phi(z) = \sum_{n \in \mathbb{Z}} e^{\pi i z n^2} = \Theta(z/2)$. Let χ be the unique nontrivial character of $\mathfrak{G}(2)/\Gamma(2)$ (which has 2 elements). Show that $\phi^4 \in M_2(\mathfrak{G}(2), \chi)$.

15. Let $f \in M_k(N, \chi)$, and set $\alpha_N = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix}$.
(a) Prove that $f|[\alpha_N]_k \in M_k(N, \bar{\chi})$, and that the map $f \mapsto f|[\alpha_N]_k$ is an isomorphism of vector spaces from $M_k(N, \chi)$ to $M_k(N, \bar{\chi})$. Prove that the square of this map

(i.e., where one uses it to go from $M_k(N, \chi)$ to $M_k(N, \bar{\chi})$ and then again to go from $M_k(N, \bar{\chi})$ back to $M_k(N, \chi)$) is the map $(-1)^k$ on $M_k(N, \chi)$.

(b) If $\chi = \bar{\chi}$, i.e., if χ takes only the values ± 1, then prove that $M_k(N, \chi) = M_k^+(N, \chi) \oplus M_k^-(N, \chi)$, where $M_k^{\pm}(N, \chi) \underset{\text{def}}{=} \{f \in M_k(N, \chi) | f|[\alpha_N]_k = \pm i^{-k} f\}$. In other words, any modular form in $M_k(N, \chi)$ can be written as a sum of one which is fixed under $i^k[\alpha_N]_k$ and one which is taken to its negative by $i^k[\alpha_N]_k$.

(c) Let $N = 4$. In Problem 17(d) below, we shall see that Θ^4 and F span $M_2(\Gamma_0(4)) = M_2(4, 1)$ (where $F = \Sigma_{n \text{ odd}} \sigma_1(n) q^n$ as in Problem 10, and 1 denotes the trivial character in $M_2(4, 1)$). Assuming this, find the matrix of $[\alpha_4]_2$ in the basis Θ^4, F; show that $M_2^+(4, 1)$ and $M_2^-(4, 1)$ are each one-dimensional; and find a basis for $M_2(\Gamma_0(4))$ consisting of eigenforms for $[\alpha_4]_2$. If you normalize these eigenvectors by requiring the coefficient of q in their q-expansions to be 1, then they are uniquely determined.

16. (a) For $k \geq 2$ even, let $f(z) = -\frac{B_k}{2k} E_k = -\frac{B_k}{2k} + \Sigma_{n=1}^{\infty} \sigma_{k-1}(n) q^n$. Express $L_f(s)$ (see the definition in (3.33)) in terms of the Riemann zeta-function.

(b) Write an Euler product for $L_f(s)$.

(c) Let $f_\chi(z) = \Sigma_{n=1}^{\infty} \sigma_{k-1}(n) \chi(n) q^n$ for a Dirichlet character χ. Write an Euler product for $L_{f_\chi}(s)$.

17. Let $F(2)$ be the fundamental domain for $\Gamma(2)$ constructed in §1 (see Fig. III.3). Then $F' = \alpha F(2)$, where $\alpha = \begin{pmatrix} 0 & -1 \\ 2 & 0 \end{pmatrix}$, is a fundamental domain for $\Gamma_0(4) = \alpha \Gamma(2) \alpha^{-1}$ (see Problem 10 in §III.1). The boundary of F' consists of: two vertical lines extending from $(-3 + i\sqrt{3})/4$ and from $(1 + i\sqrt{3})/4$ to infinity; two arcs of circles of radius $\frac{1}{2}$, one centered at $-\frac{1}{2}$ and one centered at 0; the arc of the circle of radius $\frac{1}{6}$ and center $\frac{1}{6}$ which extends from 0 to $(9 + i\sqrt{3})/28$; and the arc of the circle of radius $\frac{1}{10}$ and center $\frac{4}{10}$ which extends from $(9 + i\sqrt{3})/28$ to $\frac{1}{2}$. Consider $\Gamma_0(4)$-equivalent points on the boundary of F' to be identified.

(a) Find all elliptic points in F' (i.e., points which are Γ-equivalent to i or $\omega = (-1 + i\sqrt{3})/2$). Which are on the boundary and which are in the interior of F'?

(b) Let $f(z)$ be a nonzero modular function of weight k ($k \in \mathbb{Z}$ even) for $\Gamma_0(4)$. Let $v_p(f)$ denote the order of zero or pole of $f(z)$ at the point P. At a cusp $P = \alpha^{-1}\infty$, we define $v_p(f)$ to be the first power of q_h with nonzero coefficient in the Fourier expansion of $f|[\alpha^{-1}]_k$ (where h is the ramification index; see Problem 2 above). Prove that: $\Sigma_{p \in F'} v_p(f) = k/2$, where the summation is over all points in the fundamental domain F', including the three cusps, but taking only one point in a set of $\Gamma_0(4)$-equivalent boundary points (e.g., $\{-\frac{3}{4} + iy, \frac{1}{4} + iy\}$ or $\{-\frac{3}{4} + i\frac{\sqrt{3}}{4}, \frac{1}{4} + i\frac{\sqrt{3}}{4}, \frac{9}{28} + i\frac{\sqrt{3}}{28}\}$).

(c) Describe the zeros of $\Theta(z)^4$ and $F(z)$ (see Problems 10–11 above).

(d) Prove that Θ^4 and F span $M_2(\Gamma_0(4))$.

(e) Prove that $M_k(\Gamma_0(4)) = 0$ if $k < 0$, and it contains only the constants if $k = 0$.

(f) Prove that for $k = 2k_0$ a nonnegative even integer, any $f \in M_k(\Gamma_0(4))$ can be written as a homogeneous polynomial of degree k_0 in F and Θ^4.

(g) Prove that $S_6(\Gamma_0(4))$ is one-dimensional and is spanned by $\Theta^8 F - 16\Theta^4 F^2$.

(h) Prove that $\eta^{12}(2z) \notin S_6(\Gamma_0(2))$, but that $\eta^{12}(2z) \in S_6(\Gamma_0(4))$. Then conclude that $\eta^{12}(2z) = \Theta^8 F - 16\Theta^4 F^2$.

(i) Prove that for $k = 2k_0 \geq 6$, any $f \in S_k(\Gamma_0(4))$ can be written as a homogeneous polynomial of degree k_0 in F and Θ^4 that is divisible by $\Theta^4 F(\Theta^4 - 16F)$.

18. (a) If $f \in M_{k_1}(N, \chi_1)$ and $g \in M_{k_2}(N, \chi_2)$, show that $fg \in M_{k_1+k_2}(N, \chi_1\chi_2)$.

(b) Let χ be the unique nontrivial character of $(\mathbb{Z}/4\mathbb{Z})^*$. Show that any element of $M_1(4, \chi)$ is a constant multiple of Θ^2.
(c) With χ as in part (b), find a formula for $\dim S_k(\Gamma_1(4)) = \dim S_k(4, \chi^k)$.
(d) Let $f(z) = (\eta(z)\eta(2z))^8$, and let $g(z) = f(2z)$. Show that $S_8(\Gamma_1(4))$ is spanned by f and g.

19. Let $\Gamma' \subset \Gamma$ be a congruence subgroup, with $\bar{\Gamma} = \bigcup \alpha_j \bar{\Gamma}'$, so that $F' = \bigcup \alpha_j^{-1} F$ is a fundamental domain for Γ'. Let $f_j = f|[\alpha_j^{-1}]_k$ for $f \in M_k(\Gamma')$. Suppose that $f \in S_k(\Gamma')$, so that $f_j(z) = \Sigma_{n=1}^{\infty} a_{n,j} q_{h_j}^n$, where h_j is the ramification index of Γ' at the cusp $s_j = \alpha_j^{-1}\infty$. In particular, $f(z) = \Sigma_{n=1}^{\infty} a_n q_h^n$ at the cusp ∞.
(a) Show that there exists a constant C independent of j and x such that
$$|f_j(x + iy)| \leq Ce^{-2\pi y/h_j} \quad \text{for} \quad y > \varepsilon.$$
(b) Let $g(z) = (\operatorname{Im} z)^{k/2}|f(z)|$. Show that $g(\gamma z) = (\operatorname{Im} z)^{k/2}|f(z)|[\gamma]_k|$ for $\gamma \in \Gamma$.
(c) Show that $g_j(z) \underset{\text{def}}{=} (\operatorname{Im} z)^{k/2}|f_j(z)|$ is bounded on F.
(d) Show that $g(z)$ is bounded on F'.
(e) Show that $g(z)$ is bounded on H.
(f) Show that for any fixed y:
$$a_n = \frac{1}{h}\int_0^h f(x + iy)e^{-2\pi in(x+iy)/h}dx.$$
(g) Show that there exists a constant C_1 such that for all y:
$$|a_n| \leq C_1 y^{-k/2} e^{2\pi ny/h}.$$
(h) Choosing $y = 1/n$ in part (g), show that for $C_2 = C_1 e^{2\pi/h}$: $|a_n| \leq C_2 n^{k/2}$.
(i) Show that $|a_{n,j}|n^{-k/2}$ is similarly bounded for each j.

§4. Transformation formula for the theta-function

We first define some notation. Let d be an *odd* integer, and let c be any integer. The quadratic residue symbol $(\frac{c}{d})$ is defined in the usual way when d is a (positive) prime number, i.e., it equals 0 if $d|c$, 1 if c is a nonzero quadratic residue modulo d, and -1 otherwise. We extend this definition to arbitrary odd d as follows. First, if g.c.d.$(c, d) > 1$, then always $(\frac{c}{d}) = 0$. Next, if d is positive, we write d as a product of primes $d = \Pi_j p_j$ (not necessarily distinct), and define $(\frac{c}{d}) = \Pi_j (\frac{c}{p_j})$. If $d = \pm 1$ and $c = 0$, we adopt the convention that $(\frac{0}{\pm 1}) = 1$. Finally, if d is negative, then we define $(\frac{c}{d}) = (\frac{c}{|d|})$ if $c > 0$ and $(\frac{c}{d}) = -(\frac{c}{|d|})$ if $c < 0$.

It is easy to check that this quadratic residue symbol is bimultiplicative, i.e., it is multiplicative in c if d is held fixed and multiplicative in d if c is held fixed. It is also periodic with period d when d is positive: $(\frac{c+d}{d}) = (\frac{c}{d})$ if $d > 0$. However, one must be careful, because periodicity fails when d is negative and $c + d$ and c have different signs: $(\frac{c+d}{d}) = -(\frac{c}{d})$ if $c > 0 > c + d$. This is because of our convention that $(\frac{c}{d}) = -(\frac{c}{|d|})$ when both c and d are negative. On the other hand, this convention ensures that the usual formula $(\frac{-1}{d}) = (-1)^{(d-1)/2}$ holds whether d is positive or negative.

Next, we adopt the convention that $\sqrt{z}$ for $z \in \mathbb{C}$ always denotes the branch whose argument is in the interval $(-\pi/2, \pi/2]$. We next define ε_d for d odd by: $\varepsilon_d = \sqrt{(\frac{-1}{d})}$, i.e.,

$$\varepsilon_d = \begin{cases} 1 & \text{if } d \equiv 1 \bmod 4; \\ i & \text{if } d \equiv 3 \bmod 4. \end{cases} \tag{4.1}$$

With these definitions we finally define the "automorphy factor" $j(\gamma, z)$, which depends on $\gamma = (\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}) \in \Gamma_0(4)$ and $z \in H$:

$$j(\gamma, z) \underset{\text{def}}{=} \left(\frac{c}{d}\right)\varepsilon_d^{-1}\sqrt{cz+d} \quad \text{for} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(4), \qquad z \in H. \tag{4.2}$$

Recall our definition of the theta-function $\Theta(z) = \Sigma_{n \in \mathbb{Z}} q^{n^2} = \Sigma_{n \in \mathbb{Z}} e^{2\pi i z n^2}$.

The purpose of this section is to prove the following theorem, following Hecke [1944].

Theorem. *For* $\gamma \in \Gamma_0(4)$ *and* $z \in H$

$$\Theta(\gamma z) = j(\gamma, z)\Theta(z), \tag{4.3}$$

where $j(\gamma, z)$ *is defined by* (4.2).

Notice that the square of $j(\gamma, z)$ is $(\frac{-1}{d})(cz + d)$, and so the square of the equality is precisely what we proved in Proposition 30. Thus, for fixed $\gamma \in \Gamma_0(4)$ the ratio of the two sides of (4.3) is a holomorphic function of $z \in H$ whose square is identically 1. Thus, the ratio itself is ± 1. The content of the theorem is that this ratio is $+1$, i.e., that $j(\gamma, z)$ has the right sign.

Simple as that sounds, the theorem is by no means trivial to prove. At first, it might seem sensible to proceed as in the proof of Proposition 30, proving that (4.3) holds for generators of $\Gamma_0(4)$. However, then we would have to show that the expression $j(\gamma, z)$ in (4.2) has a certain multiplicative property which ensures that, if (4.3) holds for γ_1 and γ_2, then it must hold for $\gamma_1\gamma_2$. But that is a mess to try to show directly. We shall, in fact, conclude such a property for $j(\gamma, z)$ as a consequence of the theorem (see Problem 3 at the end of this section).

In proving the theorem, it turns out to be easier to work with the function $\phi(z) \underset{\text{def}}{=} \Theta(z/2) = \Sigma_{n \in \mathbb{Z}} e^{\pi i n^2 z}$, which we encountered in Problem 14 of §III.3. This function satisfies: $\phi(T^2 z) = \phi(z)$ (obvious from the definition) and $\phi(Sz) = \sqrt{-iz}\,\phi(z)$ (immediate from (3.4)). Hence, $\phi(\gamma z)$ has a transformation rule for any γ in the group $\mathfrak{G}(2)$ generated by $\pm T^2$, S. It is because $\mathfrak{G}(2)$ is such a large group—having only index 3 in Γ—that it is sometimes easier to work with ϕ. The corresponding group under which $\Theta(z) = \phi(\alpha z)$, $\alpha = (\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix})$, has a transformation rule is $\alpha^{-1}\mathfrak{G}(2)\alpha$ (see Problem 1 of §III.3). But $\alpha^{-1}\mathfrak{G}(2)\alpha$ is not contained in $\Gamma = SL_2(\mathbb{Z})$; its intersection with Γ is the subgroup $\Gamma_0(4)$ of index six in Γ. Thus, we can work with a "larger" subgroup of Γ (i.e., its index is smaller) if we work with $\phi(z)$ rather than $\Theta(z)$. The next lemma gives an equivalent form of the theorem in terms of $\phi(z)$.

Lemma 1. *The theorem follows if we prove the following transformation formula for* $\phi(z)$:

$$\phi(\gamma z) = i^{(1-c)/2}\left(\frac{d}{c}\right)\sqrt{-i(cz+d)}\phi(z) \tag{4.4}$$

for $\gamma \in \Gamma$ *such that* $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \pmod 2$ *with* $d \neq 0$.

PROOF. Suppose that we have the transformation rule for $\phi(z)$. We must show that Θ satisfies (4.3) for all $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(4)$. We first note that if $c = 0$, then (4.3) holds trivially, since in that case $\Theta(\gamma z) = \Theta(z)$, while $j(\gamma, z) = \varepsilon_d^{-1}\sqrt{d} = 1$ (since $d = \pm 1$). So in what follows we suppose that $c \neq 0$.

For any $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(4)$ with $c \neq 0$ we write

$$\Theta(\gamma z) = \phi\left(2\frac{az+b}{cz+d}\right) = \phi\left(\frac{2b(-1/2z) - a}{d(-1/2z) - c/2}\right) = \phi(\gamma'(-1/2z)),$$

where $\gamma' = \left(\begin{smallmatrix} 2b & -a \\ d & -c/2 \end{smallmatrix}\right)$ and $\gamma' \equiv \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \pmod 2$, because $4|c$. We apply (4.4) with γ' in place of γ and $-1/2z$ in place of z. Using the fact that $\left(\frac{-c/2}{d}\right) = \left(\frac{-2}{d}\right)\left(\frac{c}{d}\right)$, we obtain

$$\Theta(\gamma z) = i^{(1-d)/2}\left(\frac{-2}{d}\right)\left(\frac{c}{d}\right)\sqrt{-i(d(-1/2z) - c/2)}\phi(-1/2z).$$

Next, we have $\phi(-1/2z) = \Theta(-1/4z) = \sqrt{-2iz}\Theta(z)$ by (3.4). The product of the two square root terms is $\pm\sqrt{cz+d}$ (note that, because of our convention on the branch of the square root, we have $\sqrt{x}\sqrt{y} = \pm\sqrt{xy}$; for example, $\sqrt{-1}\sqrt{-1} = -\sqrt{1}$). But since the three functions $\sqrt{-i(d(-1/2z) - c/2)}$, $\sqrt{-2iz}$, and $\sqrt{cz+d}$ are all holomorphic on H, the $\pm$ must be the same for all z; so it suffices to check for any one value of z, say $z = i$. But in that case $\sqrt{-2iz} = \sqrt{2}$, and $\sqrt{x}\sqrt{y} = +\sqrt{xy}$ always holds when y is positive real. Thus, the product of the two square root terms is $\sqrt{cz+d}$, and we have

$$\Theta(\gamma z) = i^{(1-d)/2}\left(\frac{-2}{d}\right)\left(\frac{c}{d}\right)\sqrt{cz+d}\,\Theta(z).$$

To complete the proof of Lemma 1, it remains to check that $i^{(1-d)/2}\left(\frac{-2}{d}\right) = \varepsilon_d^{-1}$, which we easily do by considering the cases $d \equiv 1, 3, 5, 7 \pmod 8$. □

The remainder of this section is devoted to proving (4.4).

For a fixed odd prime p, let us denote

$$\psi(z) = \eta^p(z)/\eta(pz), \tag{4.5}$$

where $\eta(z)$ is the Dedekind eta-function, as in §2 and §3. According to Propositions 26–27, we have $\psi^3 \in M_{3(p-1)/2}(p, (\frac{\cdot}{p}))$, i.e., ψ^3 is a modular form for $\Gamma_0(p)$ with character $\chi(d) = \left(\frac{d}{p}\right)$. This is the basic tool which will be used to prove (4.4). But it will take several lemmas to relate ψ^3 and ϕ.

Lemma 2.

$$\phi(pz)/\phi^p(z) = \psi(z)/\psi^2\left(\frac{z+1}{2}\right). \tag{4.6}$$

PROOF. By Problem 11(e) in the preceding section, with z replaced by $\frac{z}{2}$ and by $\frac{pz}{2}$, we find that the left side of (4.6) is equal to

$$\frac{e^{-2\pi i/24}\eta^2(\frac{pz+1}{2})/\eta(pz)}{(e^{-2\pi i/24}\eta^2(\frac{z+1}{2})/\eta(z))^p} = e^{((p-1)/24)2\pi i}\frac{\psi(z)}{\psi^2(\frac{z+1}{2})}\cdot\frac{\eta^2(\frac{pz+1}{2})}{\eta^2(p\frac{z+1}{2})}.$$

But since $\eta(z+1) = e^{2\pi i/24}\eta(z)$, and $p((z+1)/2) = (pz+1)/2 + \frac{p-1}{2}$, it follows that the last term on the right is $e^{-(p-1)2\pi i/24}$. This proves (4.6). □

Lemma 3. *Let p be an odd prime, let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathfrak{G}(2) \cap \Gamma_0(p)$, and let $f(z) = \psi^3(\frac{z+1}{2})$. Then $f|[\gamma]_{3(p-1)/2} = (\frac{d}{p})f$.*

PROOF. Let $\alpha = \left(\begin{smallmatrix} 1 & 1 \\ 0 & 2 \end{smallmatrix}\right)$, so that $\psi((z+1)/2) = \psi(\alpha z)$. Then $\psi((\gamma z+1)/2) = \psi(\alpha\gamma z) = \psi((\alpha\gamma\alpha^{-1})\alpha z)$. Now for γ as in the lemma, we have

$$\alpha\gamma\alpha^{-1} = \begin{pmatrix} a+c & (b+d-a-c)/2 \\ 2c & d-c \end{pmatrix} \in \Gamma_0(p).$$

(Note that $b+d-a-c$ is divisible by 2 because $\gamma \in \mathfrak{G}(2)$.) Hence, by Propositions 26–27, we have

$$\psi^3\left(\frac{\gamma z+1}{2}\right) = \left(\frac{d-c}{p}\right)(2c\alpha z + d - c)^{3(p-1)/2}\psi^3(\alpha z)$$
$$= \left(\frac{d}{p}\right)(cz+d)^{3(p-1)/2}\psi^3\left(\frac{z+1}{2}\right),$$

because $d - c \equiv d \pmod p$. This is the relation asserted in the lemma. □

Lemma 4. *Let p be an odd prime, let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathfrak{G}(2) \cap \Gamma_0(p)$, and let $g(z) = \phi(pz)/\phi^p(z)$. Then $g|[\gamma]_{(1-p)/2} = (\frac{d}{p})g$.*

PROOF. We first claim that g^8 transforms trivially under γ. Let $\alpha = \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)$, and let $\gamma' = \alpha\gamma\alpha^{-1}$. Then both γ and γ' are in $\mathfrak{G}(2)$, and we can use Problem 14 of the preceding section to compute

$$g^8(z)|[\gamma]_{4(1-p)} = (cz+d)^{4(p-1)}\phi^8(\gamma'\alpha z)/\phi^{8p}(\gamma z)$$
$$= \frac{(\frac{c}{p}\alpha z + d)^{-4}\phi^8(\gamma'\alpha z)}{(cz+d)^{-4p}\phi^{8p}(\gamma z)}$$
$$= \frac{\phi^8(\alpha z)}{\phi^{8p}(z)} = g^8(z),$$

as claimed. Meanwhile, the *ninth* power of g transforms under γ by $(\frac{d}{p})$, as we see by raising both sides of (4.6) to the 9th power and using Propositions 26–27 and Lemma 3 (here $f(z)$ is as in Lemma 3):

$$g^9|[\gamma]_{9(1-p)/2} = (\psi^9/f^6)|[\gamma]_{9(1-p)/2} = \frac{(\psi^3|[\gamma]_{3(p-1)/2})^3}{(f|[\gamma]_{3(p-1)/2})^6} = \frac{((\frac{d}{p})\psi^3)^3}{((\frac{d}{p})f)^6} = \left(\frac{d}{p}\right)g^9.$$

Taking the quotient of these two relationships gives

$$g|[\gamma]_{(1-p)/2} = g^9|[\gamma]_{9(1-p)/2}/g^8|[\gamma]_{4(1-p)} = \left(\frac{d}{p}\right)g. \qquad \square$$

The next lemma generalizes Lemma 4 by replacing the prime p by an arbitrary positive odd number n.

Lemma 5. *Let n be a positive odd integer, let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \mathfrak{G}(2) \cap \Gamma_0(n)$, and let $g(z) = \phi(nz)/\phi^n(z)$. Then $g|[\gamma]_{(1-n)/2} = (\frac{d}{n})g$.*

Proof. We write $n = p_1 \cdots p_r$ as a product of primes (not necessarily distinct), and we use induction on the number r of prime factors. Lemma 4 is the case $r = 1$. Now suppose we know Lemma 5 for n; we shall prove the corresponding equality for a product $n' = np$ of $r + 1$ primes. We write

$$\frac{\phi(n'z)}{\phi^{n'}(z)} = \frac{\phi(n\alpha z)}{\phi^n(\alpha z)}\left(\frac{\phi(pz)}{\phi^p(z)}\right)^n, \tag{4.7}$$

where $\alpha = \left(\begin{smallmatrix} p & 0 \\ 0 & 1 \end{smallmatrix}\right)$. For $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(n') \cap \mathfrak{G}(2)$ we have $\gamma' = \alpha\gamma\alpha^{-1} = \left(\begin{smallmatrix} a & bp \\ c/p & d \end{smallmatrix}\right) \in \Gamma_0(n) \cap \mathfrak{G}(2)$, and so, by the induction assumption,

$$\begin{aligned}\phi(n\alpha\gamma z)/\phi^n(\alpha\gamma z) = \phi(n\gamma'\alpha z)/\phi^n(\gamma'\alpha z) &= \left(\frac{d}{n}\right)\left(\frac{c}{p}\alpha z + d\right)^{(1-n)/2}\phi(n\alpha z)/\phi^n(\alpha z)\\ &= \left(\frac{d}{n}\right)(cz + d)^{(1-n)/2}\phi(n\alpha z)/\phi^n(\alpha z).\end{aligned}$$

In addition, by Lemma 4, we have

$$\phi(p\gamma z)/\phi^p(\gamma z) = \left(\frac{d}{p}\right)(cz + d)^{(1-p)/2}\phi(pz)/\phi^p(z).$$

Combining these two relations, we see that replacing z by γz in (4.7) has the effect of multiplying by

$$\begin{aligned}\left(\frac{d}{n}\right)(cz + d)^{(1-n)/2}\left(\left(\frac{d}{p}\right)(cz + d)^{(1-p)/2}\right)^n &= \left(\frac{d}{n}\right)\left(\frac{d}{p}\right)(cz + d)^{(1-np)/2}\\ &= \left(\frac{d}{n'}\right)(cz + d)^{(1-n')/2}.\end{aligned}$$

This completes the induction step, and the proof of the lemma. $\square$

We are now ready to prove (4.4). We first note that both sides of (4.4) remain unchanged if γ is replaced by $-\gamma$ (see Problem 2 below). Hence, without loss of generality we may suppose that $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) \bmod 2$ with $c > 0$.

We now apply Lemma 5 with n replaced by the positive odd integer c, obtaining:

$$\frac{\phi(c\gamma z)}{\phi^c(\gamma z)} = \left(\frac{d}{c}\right)(cz + d)^{(1-c)/2}\frac{\phi(cz)}{\phi^c(z)}. \tag{4.8}$$

The ratio that ultimately interests us is $\phi(\gamma z)/\phi(z)$. Solving for this in (4.8) gives

$$\left(\frac{\phi(\gamma z)}{\phi(z)}\right)^c = \left(\frac{d}{c}\right)(cz + d)^{(c-1)/2}\frac{\phi(c\gamma z)}{\phi(cz)}. \tag{4.9}$$

On the other hand, we have (using: $ad - 1 = bc$)

$$c\gamma z = c\frac{az + b}{cz + d} = a - \frac{1}{cz + d}.$$

Since a is even and ϕ has period 2, this means that

$$\begin{aligned} \phi(c\gamma z) &= \phi\left(-\frac{1}{cz + d}\right) = \sqrt{-i(cz + d)}\phi(cz + d) \\ &= \sqrt{-i(cz + d)}\phi(cz). \end{aligned} \tag{4.10}$$

Combining (4.9) and (4.10) gives

$$\left(\frac{\phi(\gamma z)}{\phi(z)}\right)^c = \left(\frac{d}{c}\right)(cz + d)^{(c-1)/2}\sqrt{-i(cz + d)}. \tag{4.11}$$

Meanwhile, we saw in Problem 14 of the preceding section that ϕ^8 is invariant under $[\gamma]_4$ for $\gamma \in \mathfrak{G}(2)$, i.e.,

$$\left(\frac{\phi(\gamma z)}{\phi(z)}\right)^{8k} = (cz + d)^{4k} \quad \text{for any} \quad k \in \mathbb{Z}. \tag{4.12}$$

We now raise both sides of (4.11) to the c-th power and divide by (4.12), where k is chosen so that $c^2 = 8k + 1$. (Since c is odd, of course $c^2 \equiv 1 \bmod 8$.) The result is:

$$\frac{\phi(\gamma z)}{\phi(z)} = \left(\frac{d}{c}\right)(cz + d)^{c(c-1)/2-4k}(-i(cz + d))^{(c-1)/2}\sqrt{-i(cz + d)}.$$

But $c(c - 1)/2 - 4k + (c - 1)/2 = (c^2 - 1)/2 - 4k = 0$. Hence,

$$\frac{\phi(\gamma z)}{\phi(z)} = \left(\frac{d}{c}\right)(-i)^{(c-1)/2}\sqrt{-i(cz + d)},$$

which is the transformation formula (4.4) that we wanted to prove. This concludes the proof of the main theorem as well. □

The transformation formula for the theta-function is similar to the transformation formula for a modular form of weight k if we take $k = \frac{1}{2}$, i.e., except for a power of i the "automorphy factor" is $(cz + d)^{1/2}$. In the next chapter we shall see that there is a general theory of modular forms whose weight is a half-integer, and the transformation formula for $\Theta(z)$ plays a fundamental role in describing such functions.

PROBLEMS

1. Prove that the generalized quadratic residue symbol $(\frac{c}{d})$ as defined in this section satisfies the following form of quadratic reciprocity: if c and d are both odd integers, then

$$(\tfrac{d}{c}) = \begin{cases} (-1)^{(c-1)(d-1)/4}(\frac{c}{d}) & \text{if } c \text{ or } d \text{ is positive;} \\ -(-1)^{(c-1)(d-1)/4}(\frac{c}{d}) & \text{if } c \text{ and } d \text{ are negative.} \end{cases}$$

2. Show directly that both sides of (4.4) remain unchanged if γ is replaced by $-\gamma$.
3. (a) Show directly (using (3.4)) that the theorem (4.3) holds for the generators $-I$, T, and $ST^{-4}S$ of $\Gamma_0(4)$.
   (b) Show that the theorem would follow from part (a) if one could show that
   $$j(\alpha\beta, z) = j(\alpha, \beta z)j(\beta, z) \quad \text{for all} \quad \alpha, \beta \in \Gamma_0(4). \tag{4.13}$$
   (c) Conversely, show that the theorem proved in this section implies the relation (4.13).

§5. The modular interpretation, and Hecke operators

A basic feature of modular forms is their interpretation as functions on lattices. More precisely, we consider the most important cases of a congruence subgroup Γ': $\Gamma' = \Gamma$, $\Gamma_1(N)$, $\Gamma_0(N)$ or $\Gamma(N)$. (Of course, $\Gamma = \Gamma_1(1) = \Gamma_0(1) = \Gamma(1)$, so everything we say about the cases $\Gamma_1(N)$, $\Gamma_0(N)$ or $\Gamma(N)$ will apply to Γ if we set $N = 1$.) By a "modular point" for Γ' we mean:

(i) for $\Gamma' = \Gamma$: a lattice $L \subset \mathbb{C}$;
(ii) for $\Gamma' = \Gamma_1(N)$: a pair (L, t), where L is a lattice in $\mathbb{C}$, and $t \in \mathbb{C}/L$ is a point of exact order N;
(iii) for $\Gamma' = \Gamma_0(N)$: a pair (L, S), where L is a lattice in $\mathbb{C}$, and $S \subset \mathbb{C}/L$ is a cyclic subgroup of order N, i.e., $S = \mathbb{Z}t$ for some point $t \in \mathbb{C}/L$ of exact order N.
(iv) for $\Gamma' = \Gamma(N)$: a pair $(L, \{t_1, t_2\})$, where $t_1, t_2 \in \mathbb{C}/L$ have the property that every $t \in \frac{1}{N}L/L$ is of the form $t = mt_1 + nt_2$, i.e., t_1, t_2 form a basis for the points of order N (in particular, t_1 and t_2 must each have exact order N).

Given a lattice L, in general there will be several modular points of the form (L, t), (L, S), or $(L, \{t_1, t_2\})$. However, when $N = 1$, there is only one modular point corresponding to each L, and we identify it with the modular point L for Γ.

Let $k \in \mathbb{Z}$. In each case (i)–(iv), we consider complex-valued functions F on the set of modular points which are of "weight k" in the following sense. If we scale a modular point by a nonzero complex number λ, then the value of F changes by a factor of λ^{-k}. That is, for $\lambda \in \mathbb{C}^*$ we consider $\lambda L =$

$\{\lambda l | l \in L\}$, $\lambda t \in \mathbb{C}/\lambda L$, $\lambda S = \{\lambda t | t \in S\} \subset \mathbb{C}/\lambda L$. Then F is defined to be of weight k if for all $\lambda \in \mathbb{C}^*$

case (i) $F(\lambda L) = \lambda^{-k} F(L)$ for all modular points L;
case (ii) $F(\lambda L, \lambda t) = \lambda^{-k} F(L, t)$ for all modular points (L, t);
case (iii) $F(\lambda L, \lambda S) = \lambda^{-k} F(L, S)$ for all modular points (L, S);
case (iv) $F(\lambda L, \{\lambda t_1, \lambda t_2\}) = \lambda^{-k} F(L, \{t_1, t_2\})$ for all modular points $(L, \{t_1, t_2\})$.

An example of such a function of weight k is

$$G_k(L) = \sum_{0 \neq l \in L} l^{-k} \qquad (k > 2 \text{ even}). \tag{5.1}$$

Notice that any function F in case (i), such as G_k, automatically gives a function for the other groups; for example, by setting $F(L, t) = F(L)$.

Given a function F of weight k, we define two corresponding functions $\tilde{F}$ and f as follows. $\tilde{F}(\omega)$ is a complex-valued function on column vectors $\omega = \binom{\omega_1}{\omega_2}$ such that $\omega_1/\omega_2 \in H$; $f(z)$ is a function on the upper half-plane H. Let L_ω be the lattice spanned by ω_1 and ω_2, and let L_z be the lattice spanned by z and 1. Given F as above, we define

case (i) $\tilde{F}(\omega) = F(L_\omega)$;
case (ii) $\tilde{F}(\omega) = F(L_\omega, \omega_2/N)$;
case (iii) $\tilde{F}(\omega) = F(L_\omega, \mathbb{Z}\omega_2/N)$;
case (iv) $\tilde{F}(\omega) = F(L_\omega, \{\omega_1/N, \omega_2/N\})$.

In all cases we define $f(z) = \tilde{F}\binom{z}{1}$. Thus, for example, the function $f(z)$ that corresponds to $G_k(L)$ (see (5.1)) is the Eisenstein series we denoted $G_k(z)$ in §2 (see (2.5)).

For $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma = SL_2(\mathbb{Z})$, we define the action of γ on functions of ω by the rule $\gamma\tilde{F}(\omega) = \tilde{F}(\gamma\omega)$, where $\gamma\omega$ is the usual multiplication of a column vector by a matrix.

Proposition 31. *Let $k \in \mathbb{Z}$, and let $\Gamma' = \Gamma$, $\Gamma_1(N)$, $\Gamma_0(N)$ or $\Gamma(N)$. The above association of F with $\tilde{F}$ and f gives a one-to-one correspondence between the following sets of complex-valued functions:* (1) *F on modular points which have weight k;* (2) *$\tilde{F}$ on column vectors ω which are invariant under γ for $\gamma \in \Gamma'$ and satisfy $\tilde{F}(\lambda\omega) = \lambda^{-k}\tilde{F}(\omega)$;* (3) *$f$ on H which are invariant under $[\gamma]_k$ for $\gamma \in \Gamma'$.*

Proof. We shall treat case (ii), and leave the other cases as exercises. Suppose $\gamma \in \Gamma_1(N)$ and F is a weight k function of modular points (L, t). We first compute:

$$\tilde{F}(\gamma\omega) = F\left(L_{\gamma\omega}, \frac{c\omega_1 + d\omega_2}{N}\right) = F(L_\omega, \omega_2/N) = \tilde{F}(\omega),$$

because $L_{a\omega_1 + b\omega_2, c\omega_1 + d\omega_2} = L_{\omega_1, \omega_2}$ (since $\gamma \in \Gamma$) and $(c\omega_1 + d\omega_2)/N \equiv \omega_2/N \bmod L$ (since $\gamma \in \Gamma_1(N)$). We also have

$$\tilde{F}(\lambda\omega) = F(L_{\lambda\omega}, \lambda\omega_2/N) = \lambda^{-k}F(L_\omega, \omega_2/N) = \lambda^{-k}\tilde{F}(\omega).$$

Next, we have

$$f(\gamma z) = F\left(L_{(az+b)/(cz+d)}, \frac{1}{N}\right) = (cz+d)^k F\left(L_{az+b,cz+d}, \frac{cz+d}{N}\right),$$

because F has weight k. But the lattice spanned by $az+b$ and $cz+d$ is L_z, and $(cz+d)/N \equiv \frac{1}{N} \bmod L_z$; hence

$$f(\gamma z) = (cz+d)^k F\left(L_z, \frac{1}{N}\right) = (cz+d)^k f(z).$$

Thus, the $\tilde{F}$ and f corresponding to F have the properties claimed.

To show the correspondence in the other direction, given $f(z)$ we define $\tilde{F}(\omega)$ to be $\omega_2^{-k} f(\omega_1/\omega_2)$; and given $\tilde{F}$ we define $F(L, t)$ to be $\tilde{F}(\omega)$, where ω is chosen to be any basis of L such that $\omega_2/N \equiv t \bmod L$. One must first check that the definition of F makes sense (i.e., that such a basis ω exists), and that the definition of F is independent of the choice of such a basis ω. The first point is routine, using the fact that t has exact order N in $\mathbb{C}/L$, and the second point follows immediately because any other such basis must be of the form $\gamma\omega$ with $\gamma \in \Gamma_1(N)$. It is also easy to check that, once f is invariant under $[\gamma]_k$ for $\gamma \in \Gamma_1(N)$, it follows that $\tilde{F}$ is invariant under γ and has weight k; and that, if $\tilde{F}$ has weight k, then so does the corresponding F. The construction going from F to $\tilde{F}$ to f and the construction going from f to $\tilde{F}$ to F are clearly inverse to one another. This concludes the proof. □

We say that F is a modular function/ modular form/ cusp form if the corresponding f is a modular function/ modular form/ cusp form as defined in §3.

We now discuss the Hecke operators acting on modular forms of weight k for $\Gamma_1(N)$. We could define them directly on $f(z) \in M_k(\Gamma_1(N))$. However, the definition appears more natural when given in terms of the corresponding functions F on modular points.

Let $\mathscr{L}$ denote the $\mathbb{Q}$-vector space of formal finite linear combinations of modular points, i.e., $\mathscr{L} = \oplus\mathbb{Q}e_{L,t}$ is the direct sum of infinitely many one-dimensional spaces, one for each pair (L, t), where L is any lattice in $\mathbb{C}$ and $t \in \mathbb{C}/L$ is any point of exact order N. A linear map $T\colon \mathscr{L} \to \mathscr{L}$ can be given by describing the image $Te_{L,t} = \Sigma a_n e_{P_n}$ of each basis element; here $\{P_n\}$ are a finite set of modular points.

For each positive integer n we define a linear map $T_n\colon \mathscr{L} \to \mathscr{L}$ by the following formula giving the image of the basis vector $e_{L,t}$:

$$T_n(e_{L,t}) = \frac{1}{n}\sum e_{L',t}, \tag{5.2}$$

where the summation is over all lattices L' containing L with index n such that (L', t) is a modular point. (Here for $t \in \mathbb{C}/L$ we still use the letter t to

denote the image of t modulo the larger lattice L'.) In other words, $L'/L \subset \mathbb{C}/L$ is a subgroup of order n, and t must have exact order N modulo the larger lattice L' as well as modulo L. The latter condition means that the only multiples $\mathbb{Z}t$ which are in L'/L are the multiples $\mathbb{Z}Nt$ which are in L. In the case $N = 1$, i.e., $\Gamma' = \Gamma$, this condition disappears, and we sum over all lattices L' with $[L' : L] = n$. The condition on t is also empty if n and N have no common factor. To see this, suppose that g.c.d.$(n, N) = 1$, and suppose that $N't \in L'$. Then the order of $N't$ in L'/L divides N (because $N'Nt \in L$) and divides n (because $\# L'/L = n$), and so divides g.c.d.$(N, n) = 1$. Thus $N't \in L$.

Notice that the sum in (5.2) is finite, since any lattice L' in the sum must be contained in $\frac{1}{n}L = \{\frac{1}{n}l \mid l \in L\}$, because each element of L'/L has order dividing $n = \# L'/L$. Thus, each L' in the sum corresponds to a subgroup of order n in $\frac{1}{n}L/L \approx (\mathbb{Z}/n\mathbb{Z})^2$.

Note that $T_1 = 1 =$ the identity map.

Next, for any positive integer *n prime to N* we define another linear map $T_{n,n} \colon \mathscr{L} \to \mathscr{L}$ by

$$T_{n,n}(e_{L,t}) = \frac{1}{n^2} e_{(1/n)L,t}. \tag{5.3}$$

Note that t has exact order N modulo $\frac{1}{n}L$, because g.c.d.$(N, n) = 1$. Again we are using the same letter t to denote an element in $\mathbb{C}/L$ and the corresponding element in $\mathbb{C}/\frac{1}{n}L$.

It is easy to check the commutativity of the operators

$$T_{n_1,n_1} T_{n_2,n_2} = T_{n_1 n_2, n_1 n_2} = T_{n_2,n_2} T_{n_1,n_1}; \qquad T_{n,n} T_m = T_m T_{n,n}. \tag{5.4}$$

It is also true, but not quite so trivial to prove, that the T_m's commute with one another for different m's. This will follow from the next proposition.

Proposition 32. (a) *If* g.c.d.$(m, n) = 1$, *then* $T_{mn} = T_m T_n$*; in particular,* T_m *and* T_n *commute.*

(b) *If p is a prime dividing N, then* $T_{p^l} = T_p^l$.

(c) *If p is a prime not dividing N, then for* $l \geq 2$

$$T_{p^l} = T_{p^{l-1}} T_p - p T_{p^{l-2}} T_{p,p}. \tag{5.5}$$

Proof. (a) In the sum (5.2) for T_{mn}, the L' correspond to certain subgroups S' of order mn in $\frac{1}{mn}L/L$, namely, those which have trivial intersection with the subgroup $\mathbb{Z}t \subset \mathbb{C}/L$. Since g.c.d.$(m, n) = 1$, it follows that any such S' has a unique subgroup S'' of order n; if $L'' \supset L$ is the lattice corresponding to S'', then S'/S'' gives a subgroup of order m in $\frac{1}{m}L''/L''$. Both S'' and S'/S'' have nontrivial intersection with $\mathbb{Z}t$. Conversely, given $S'' = L''/L \subset \frac{1}{n}L/L$ of order n and a subgroup $S' = L'/L'' \subset \frac{1}{m}L''/L''$ of order m, where both subgroups have trivial intersection with $\mathbb{Z}t$, we have a unique subgroup $L'/L \subset \frac{1}{mn}L/L$ of order mn with nontrivial intersection with $\mathbb{Z}t$. This shows

that the modular points that occur in $T_{mn}(e_{L,t}) = \frac{1}{mn}\Sigma\, e_{L',t}$ and in $T_m(T_n(e_{L,t})) = \frac{1}{n}\Sigma\, T_m(e_{L'',t})$ are the same.

(b) By induction, it suffices to show that $T_{p^{l-1}}T_p = T_{p^l}$ for $l \geq 2$. Let $t' = \frac{N}{p}t$. Then $T_{p^l}(e_{L,t}) = p^{-l}\Sigma\, e_{L',t}$, where the summation is over all $L' \supset L$ such that $L'/L \subset p^{-l}L/L$ has order p^l and does not contain t'. Notice that L'/L must be cyclic, since otherwise it would contain a (p, p)-subgroup of $p^{-l}L/L$. There is only one such (p, p)-subgroup, namely $\frac{1}{p}L/L$, and $t' \in \frac{1}{p}L/L$, since $pt' = Nt \in L$. Once we know that L'/L must be cyclic, we can use the same argument as in part (a). Namely, for each L' that occurs in the sum for $T_{p^l}(e_{L,t})$ there is a unique cyclic subgroup of order p in L'/L; the corresponding lattice L'' occurs in the sum for $T_p(e_{L,t})$, and L' is one of the lattices that occur in $T_{p^{l-1}}(e_{L'',t})$. This shows the equality in part (b).

(c) Since $p \nmid N$, the condition about the order of t in $\mathbb{C}/L'$ is always fulfilled. We have $T_{p^{l-1}}T_p(e_{L,t}) = p^{-l}\Sigma_{L''}\Sigma_{L'}\, e_{L',t}$, where the first summation is over all lattices L'' such that $S'' = L''/L$ has order p, and the second summation is over all L' such that $S' = L'/L''$ has order p^{l-1}. On the other hand, $T_{p^l}(e_{L,t}) = p^{-l}\Sigma_{L'}\, e_{L',t}$, where the summation is over all L' such that L'/L has order p^l. Clearly, every L' in the inner sum for $T_{p^{l-1}}T_p$ is an L' of the form in the sum for T_{p^l}, and every L' in the latter sum is an L' of the form in the former sum. But we must count how many different pairs L'', L' in the double sum lead to the same L'. First, if L'/L is cyclic, then there is only one possible L''. But if L'/L is not cyclic, i.e., if $L'/L \supset \frac{1}{p}L/L$, then L'' can be an arbitrary lattice such that L''/L has order p. Since there are $p + 1$ such lattices (for example, they are in one-to-one correspondence with the points on the projective line over the field of p elements), it follows that there are p extra times that $e_{L',t}$ occurs in the double sum for $T_{p^{l-1}}T_p$. Thus,

$$T_{p^l}(e_{L,t}) = T_{p^{l-1}}T_p(e_{L,t}) - p \cdot p^{-l} \sum_{\substack{L' \supset (1/p)L \\ [L':(1/p)L] = p^{l-2}}} e_{L',t}.$$

But

$$T_{p^{l-2}}T_{p,p}(e_{L,t}) = \frac{1}{p^2}T_{p^{l-2}}e_{(1/p)L,t} = p^{-l} \sum_{[L':(1/p)L] = p^{l-2}} e_{L',t}.$$

This concludes the proof of part (c). □

If $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$ is the prime factorization of the positive integer n, then Proposition 32(a) says that $T_n = T_{p_1^{\alpha_1}} \cdots T_{p_r^{\alpha_r}}$. Then parts (b)–(c) show that each $T_{p_j^{\alpha_j}}$ is a polynomial in T_{p_j} and T_{p_j,p_j}. It is easy to see from this and (5.4) that all of the T_n's commute with each other. Thus, the operators $T_{n,n}$ (n a positive integer prime to N) and T_m (m any positive integer) generate a commutative algebra $\mathscr{H}$ of linear maps from $\mathscr{L}$ to $\mathscr{L}$; actually, $\mathscr{H}$ is generated by the $T_{p,p}$ ($p \nmid N$ a prime) and the T_p (p any prime).

There is an elegant way to summarize the relations in Proposition 32 as formal power series identities, where the coefficients of the power series

are elements in $\mathscr{H}$. First, for $p\,|\,N$, we can restate Proposition 32(b) as follows:

$$\sum_{l=0}^{\infty} T_{p^l}X^l = \frac{1}{1 - T_pX}, \qquad p\,|\,N, \tag{5.6}$$

i.e., $(\Sigma\, T_{p^l}X^l)(1 - T_pX) = 1$ in $\mathscr{H}[[X]]$. This follows from Proposition 32(b) because, equating coefficients, we see that the coefficient of X^l is $T_{p^l} - T_{p^{l-1}}T_p$. Similarly, for $p\nmid N$, part (c) of Proposition 32 is equivalent to the identity

$$\sum_{l=0}^{\infty} T_{p^l}X^l = \frac{1}{1 - T_pX + pT_{p,p}X^2}, \qquad p\nmid N, \tag{5.7}$$

i.e., if we multiply both sides of (5.7) by $1 - T_pX + pT_{p,p}X^2$ and equate coefficients of powers of X, we see that (5.7) is equivalent to the equalities

$$T_1 = 1, \quad T_p - T_p = 0, \quad T_{p^l} - T_{p^{l-1}}T_p + pT_{p^{l-2}}T_{p,p} \quad \text{for} \quad l \geq 2.$$

To incorporate part (a) of Proposition 32, we introduce a new variable s by putting $X = p^{-s}$ for each p in (5.6) or (5.7). We then take the product of (5.6) over p with $p\,|\,N$ and (5.7) over all p with $p\nmid N$:

$$\prod_{\text{all } p}\sum_{l=0}^{\infty} T_{p^l}p^{-ls} = \prod_{p|N}\frac{1}{1 - T_pp^{-s}}\prod_{p\nmid N}\frac{1}{1 - T_pp^{-s} + T_{p,p}p^{1-2s}}.$$

But, by part (a) of the proposition, when we multiply together the sums on the left in this equality, we obtain $\Sigma\, T_nn^{-s}$, where the sum is over all positive integers n. The proof is exactly like the proof of the Euler product for the Riemann zeta-function. We use the factorization $n = p_1^{\alpha_1} \cdots p_r^{\alpha_r}$, and the relation: $T_nn^{-s} = (T_{p_1^{\alpha_1}}p_1^{-\alpha_1 s}) \cdots (T_{p_r^{\alpha_r}}p_r^{-\alpha_r s})$. We hence conclude that

$$\sum_{n=1}^{\infty} T_nn^{-s} = \prod_{p|N}\frac{1}{1 - T_pp^{-s}}\prod_{p\nmid N}\frac{1}{1 - T_pp^{-s} + T_{p,p}p^{1-2s}}. \tag{5.8}$$

For d an integer prime to N, let $[d]\colon \mathscr{L} \to \mathscr{L}$ be the linear map defined on basis elements by $[d]e_{L,t} = e_{L,dt}$. Note that dt has exact order N in $\mathbb{C}/L$ because g.c.d.$(d, N) = 1$. Also note that $[d]$ depends only on d modulo N, i.e., we have an action of the group $(\mathbb{Z}/N\mathbb{Z})^*$ on $\mathscr{L}$.

We now consider functions F on modular points and the corresponding functions $f(z)$ on H. Again we suppose that we are in the case $\Gamma' = \Gamma_1(N)$. If $T\colon \mathscr{L} \to \mathscr{L}$ is a linear map given on basis elements by equations of the form $T(e_{L,t}) = \Sigma\, a_ne_{P_n}$, then we have a corresponding linear map (which we also denote T) on the vector space of complex-valued functions on modular points: $TF(L, t) = \Sigma\, a_nF(P_n)$. For example,

$$[d]F(L, t) = F(L, dt) \qquad \text{(here g.c.d.}(d, N) = 1);$$

$$T_{n,n}F(L, t) = n^{-2}F\left(\frac{1}{n}L, t\right) \qquad (\text{g.c.d.}(n, N) = 1); \tag{5.9}$$

$$T_nF(L, t) = \frac{1}{n}\sum_{L'} F(L', t),$$

where the last summation is over all modular points (L', t) such that $[L' : L] = n$, as in (5.2).

Proposition 33. *Suppose that $F(L, t)$ corresponds to a function $f(z)$ on H which is in $M_k(\Gamma_1(N))$. Then $[d]F$, $T_{n,n}F$, and T_nF also correspond to functions (denoted $[d]f$, $T_{n,n}f$, and T_nf) in $M_k(\Gamma_1(N))$. If f is a cusp form, then so are $[d]f$, $T_{n,n}f$, and T_nf. Thus, $[d]$, $T_{n,n}$, and T_n may be regarded as linear maps on $M_k(\Gamma_1(N))$ or on $S_k(\Gamma_1(N))$. In this situation, let χ be a Dirichlet character modulo N. Then $f\in M_k(N, \chi)$ if and only if $[d]F = \chi(d)F$, i.e., if and only if*

$$F(L, dt) = \chi(d)F(L, t) \quad \text{for } d\in(\mathbb{Z}/N\mathbb{Z})^*. \tag{5.10}$$

Proof. To show that $[d]f$, $T_{n,n}f$ and T_nf are invariant under $[\gamma]_k$ for $\gamma\in\Gamma_1(N)$, by Proposition 31 it suffices to show that $[d]F(L, t)$, $T_{n,n}F(L, t)$, and $T_nF(L, t)$ have weight k, i.e., $[d]F(\lambda L, \lambda t) = \lambda^{-k}[d]F(L, t)$, $T_{n,n}F(\lambda L, \lambda t) = \lambda^{-k}T_{n,n}F(L, t)$, and $T_nF(\lambda L, \lambda t) = \lambda^{-k}T_nF(L, t)$; but this all follows trivially from the definitions. We next check the condition at the cusps.

Note that if $\alpha = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)\in GL_2^+(\mathbb{Q})$ and if $f(z)$ corresponds to $F(L, t)$, then

$$\begin{aligned} f(z)|[\alpha]_k &= (\det \alpha)^{k/2}(cz + d)^{-k}F\left(L_{\alpha z}, \frac{1}{N}\right) \\ &= (\det \alpha)^{k/2}F\left(L_{az+b,cz+d}, \frac{cz+d}{N}\right). \end{aligned}$$

In particular, if $\alpha\in\Gamma$, then this equals $F(L_z, (cz + d)/N)$. Next, for each $d\in(\mathbb{Z}/N\mathbb{Z})^*$, choose a fixed $\sigma_d\in\Gamma$ such that $\sigma_d \equiv \left(\begin{smallmatrix} 1/d & 0 \\ 0 & d\end{smallmatrix}\right) \bmod N$. (This is possible, because g.c.d.$(d, N) = 1$, and the map $\Gamma \to SL_2(\mathbb{Z}/N\mathbb{Z})$ is surjective, by Problem 2 of §III.1.) We then have

$$f(z)|[\sigma_d]_k = F\left(L_z, \frac{d}{N}\right) = [d]F\left(L_z, \frac{1}{N}\right) = [d]f(z),$$

i.e., $[d]f = f|[\sigma_d]_k$. Thus, for $\gamma_0\in\Gamma$ we check (3.8) for $[d]f$ as follows: $[d]f|[\gamma_0]_k = f|[\sigma_d\gamma_0]_k$, which has a q-expansion of the required type, i.e., $[d]f$ satisfies (3.8) if f does. Similarly, we find that $T_{n,n}f(z) = n^{-2}F(\frac{1}{n}L_z, \frac{1}{N}) = n^{k-2}F(L_z, \frac{n}{N}) = n^{k-2}[n]f(z)$, so this case has already been covered.

We next consider the cusp condition for $T_nf(z)$, which is a sum of functions of the form $\frac{1}{n}F(L', \frac{1}{N})$, where L' is a lattice containing L_z with index n. We take such an L' and let (ω_1, ω_2) be a basis for L'. Since $L_z \subset L'$ with index n, there is a matrix τ with integral entries and determinant n such that $\binom{z}{1} = \tau\omega$ (ω denotes the column vector with entries ω_1 and ω_2). We can choose a set T of such τ (independently of z) such that

$$T_nf(z) = \frac{1}{n}\sum_{\tau\in T} F\left(L_{\tau^{-1}\binom{z}{1}}, \frac{1}{N}\right).$$

We now consider each $F(L_\omega, \frac{1}{N})$, where $\binom{z}{1} = \tau\binom{\omega_1}{\omega_2}$. We find a $\gamma\in\Gamma$ such that $\gamma\tau^{-1} = \frac{1}{n}\left(\begin{smallmatrix} a & b \\ 0 & d\end{smallmatrix}\right)$ with zero lower-left entry and a, b, d integers with $ad = n$

(it is an easy exercise to see that this can be done). The lattice spanned by $\binom{\omega_1}{\omega_2} = \tau^{-1}\binom{z}{1}$ is the same as that spanned by $\gamma\binom{\omega_1}{\omega_2} = \frac{1}{n}\begin{pmatrix} a & b \\ 0 & d\end{pmatrix}\binom{z}{1}$. Thus,

$$F\left(L_\omega, \frac{1}{N}\right) = F\left(L_{(az+b)/n, d/n}, \frac{1}{N}\right) = a^k F\left(L_{(az+b)/d, 1}, \frac{a}{N}\right)$$
$$= a^k[a]f((az+b)/d).$$

But if $f(z)$ satisfies (3.8), then so does $f((az+b)/d)$, and $[a]$ also preserves (3.8), as shown above. This proves the cusp condition for $T_n f$.

Finally, let $\gamma_d \in \Gamma_0(N)$ be any element with lower-right entry d. As shown above, $f(z)|[\gamma_d]_k = F(L_z, \frac{d}{N}) = [d]F(L_z, \frac{1}{N})$. Thus, $f|[\gamma_d]_k$ corresponds to $[d]F$, and so $f|[\gamma_d]_k = \chi(d)f$ if and only if $[d]F = \chi(d)F$. This completes the proof of the proposition. □

We saw before (Proposition 28) that a function $f \in M_k(\Gamma_1(N))$ can be written as a sum of functions in $M_k(N, \chi)$ for different Dirichlet characters χ. Thus, using the one-to-one correspondence in Proposition 31, we can write a modular form $F(L, t)$ as a direct sum of F's which satisfy (5.10) for various χ.

Proposition 34. *The operators T_n and $T_{n,n}$ commute with $[d]$, and preserve the space of $F(L, t)$ of weight k which satisfy* (5.10). *If $F(L, t)$ has weight k and satisfies* (5.10), *then $T_{n,n}F = n^{k-2}\chi(n)F$.*

Proof. That the operators commute follows directly from the definitions. Next, if $[d]F = \chi(d)F$, it follows that $[d]T_nF = T_n[d]F = \chi(d)T_nF$ and $[d]T_{n,n}F = T_{n,n}[d]F = \chi(d)T_{n,n}F$. This is just the simple fact from linear algebra that the eigenspace for an operator $[d]$ with a given eigenvalue is preserved under any operator which commutes with $[d]$. Finally, if $F(L, t)$ satisfies (5.10), then $T_{n,n}F(L, t) = n^{-2}F(\frac{1}{n}L, t) = n^{k-2}F(L, nt) = n^{k-2}[n]F(L, t) = n^{k-2}\chi(n)F(L, t)$. □

If we translate the action of T_n, $T_{n,n}$ and $[d]$ from functions $F(L, t)$ to functions $f(z)$ on H, then Proposition 34 becomes

Proposition 35. *T_n and $T_{n,n}$ preserve $M_k(N, \chi)$, and also $S_k(N, \chi)$. For $f \in M_k(N, \chi)$ the action of $T_{n,n}$ is given by $T_{n,n}f = n^{k-2}\chi(n)f$.*

Proposition 36. *The operators T_n on $M_k(N, \chi)$ satisfy the formal power series identity*

$$\sum_{n=1}^{\infty} T_n n^{-s} = \prod_{\text{all } p} (1 - T_p p^{-s} + \chi(p)p^{k-1-2s})^{-1}. \tag{5.11}$$

Proof. We simply use (5.8) and observe that for $p \nmid N$ we have $T_{p,p}f = p^{k-2}\chi(p)f$, while for $p|N$ the term on the right in (5.11) becomes $(1 - T_p p^{-s})^{-1}$ because $\chi(p) = 0$. □

As a special case of these propositions, suppose we take χ to be the trivial character of $(\mathbb{Z}/N\mathbb{Z})^*$. Then $M_k(N, \chi) = M_k(\Gamma_0(N))$, and the modular forms $f \in M_k(N, \chi)$ correspond to $F(L, t)$ on which $[d]$ acts trivially, i.e., $F(L, t) = F(L, dt)$ for all $d \in (\mathbb{Z}/N\mathbb{Z})^*$. Such $F(L, t)$ are in one-to-one correspondence with functions $F(L, S)$, where S is a cyclic subgroup of order N in $\mathbb{C}/L$. Namely, choose t to be any generator of S, and set $F(L, S) = F(L, t)$. The fact that $F(L, t) = F(L, dt)$ means that it makes no difference which generator of S is chosen. Conversely, given $F(L, S)$, define $F(L, t) = F(L, S_t)$, where $S_t = \mathbb{Z}t$ is the subgroup of $\mathbb{C}/L$ generated by t. So we have just verified that functions of modular points in the sense of case (iii) at the beginning of this section correspond to modular forms for $\Gamma_0(N)$.

We now examine the effect of the Hecke operator T_m on the q-expansion at ∞ of a modular form $f(z) \in M_k(N, \chi)$. That is, if we write $f(z) = \Sigma\, a_n q^n$ and $T_m f(z) = \Sigma\, b_n q^n$, $q = e^{2\pi i z}$, we want to express b_n in terms of the a_n.

We first introduce some notation. If $f \in \mathbb{C}[[q]]$, $f = \Sigma\, a_n q^n$, we define

$$V_m f = \sum a_n q^{mn}; \qquad U_m f = \sum a_n q^{n/m}, \tag{5.12}$$

where the latter summation is only over n divisible by m. Note that $U_1 = V_1 =$ identity, and $U_m \circ V_m$ is the identity, while $V_m \circ U_m$ is the map on power series which deletes all terms with n not divisible by m. Suppose that $f(z) = \Sigma\, a_n q^n$, $q = e^{2\pi i z}$, converges for $z \in H$. Then we clearly have:

$$V_m f(z) = f(mz); \qquad U_m f(z) = \frac{1}{m}\sum_{j=0}^{m-1} f\left(\frac{z+j}{m}\right). \tag{5.13}$$

Proposition 37. *Let $f(z) = \Sigma_{n=0}^{\infty} a_n q^n$, $q = e^{2\pi i z}$, $f \in M_k(N, \chi)$, and let $T_p f(z) = \Sigma_{n=0}^{\infty} b_n q^n$. Then*

$$b_n = a_{pn} + \chi(p) p^{k-1} a_{n/p}, \tag{5.14}$$

where we take $\chi(p) = 0$ if $p \mid N$ and we take $a_{n/p} = 0$ if n is not divisible by p. In other words,

$$T_p = U_p + \chi(p) p^{k-1} V_p \quad \text{on} \quad M_k(N, \chi). \tag{5.15}$$

Proof. We have $T_p f(z) = \frac{1}{p}\Sigma_{L'} F(L', \frac{1}{N})$, where F is the function on modular points which corresponds to f and the sum is over all lattices L' containing L_z with index p such that $\frac{1}{N}$ has order N modulo L'. Such L' are contained in the lattice $\frac{1}{p}L_z$ generated by $\frac{z}{p}$ and $\frac{1}{p}$, and the lattices of index p are in one-to-one correspondence with the projective line $\mathbb{P}^1_{\mathbb{F}_p}$ over the field of p elements $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$. Namely, the point in $\mathbb{P}^1_{\mathbb{F}_p}$ with homogeneous coordinates (a_1, a_2) corresponds to the lattice generated by L_z and $(a_1 z + a_2)/p$. Thus, there are $p + 1$ possible L' corresponding to $(1, j)$ for $j = 0, 1, \ldots, p-1$ and $(0, 1)$. If $p \nmid N$, then all $p + 1$ of these lattices L' are included; if $p \mid N$, then the last lattice (generated by L_z and $\frac{1}{p}$) must be omitted, since $\frac{1}{N}$ has order $\frac{N}{p}$ in that case. Note that the lattice generated by L_z and $(z + j)/p$ is $L_{(z+j)/p}$. Thus, if $p \mid N$ we have $T_p f(z) = \frac{1}{p}\Sigma_{j=0}^{p-1} F(L_{(z+j)/p}, \frac{1}{N}) = \frac{1}{p}\Sigma_{j=0}^{p-1} f(\frac{z+j}{p}) = U_p f(z)$. If

$p \nmid N$, then we have the same sum plus one additional term corresponding to the lattice generated by L_z and $\frac{1}{p}$; this lattice is $\frac{1}{p}L_{pz}$. Thus, in that case

$$T_p f(z) = U_p f(z) + \frac{1}{p} F\left(\frac{1}{p} L_{pz}, \frac{1}{N}\right) = U_p f(z) + p^{k-1} F\left(L_{pz}, \frac{p}{N}\right)$$
$$= U_p f(z) + p^{k-1}\chi(p) F\left(L_{pz}, \frac{1}{N}\right) = U_p f(z) + p^{k-1}\chi(p) f(pz),$$

which is (5.15). The expression (5.14) for the b_n follows directly from (5.15) if we use the expressions (5.12) for the operators V_p and U_p. □

As a consequence of Proposition 37 we have the factorization

$$1 - T_p X + \chi(p)p^{k-1}X^2 = (1 - U_p X)(1 - \chi(p)p^{k-1}V_p X), \tag{5.16}$$

where both sides are regarded as polynomials in the variable X whose coefficients are in the algebra of operators on the subspace of $\mathbb{C}[[q]]$ formed by the q-expansions of elements $f(z) \in M_k(N, \chi)$. To see (5.16), note that the equality of coefficients of X is precisely (5.15), while the coefficients of X^2 agree because $U_p \circ V_p = 1$.

Proposition 38. *We have the following formal identity:*

$$\sum_{n=1}^{\infty} T_n n^{-s} = \left(\sum_{n=1}^{\infty} \chi(n)n^{k-1}V_n n^{-s}\right)\left(\sum_{n=1}^{\infty} U_n n^{-s}\right), \tag{5.17}$$

or, equivalently,

$$T_n = \sum_{d|n} \chi(d)d^{k-1} V_d \circ U_{n/d}. \tag{5.18}$$

PROOF. By (5.11) and (5.16) we find that the left side of (5.17) is equal to

$$\prod_p ((1 - U_p p^{-s})(1 - \chi(p)p^{k-1}V_p p^{-s}))^{-1}.$$

Since the U_p and V_p do not commute, we must be careful about the order of the factors. Moving the inverse operation inside the outer parentheses, we reverse the order, obtaining

$$\prod_p ((1 - \chi(p)p^{k-1}V_p p^{-s})^{-1}(1 - U_p p^{-s})^{-1}).$$

Note that U_{p_1} and V_{p_2} *do* commute for $p_1 \neq p_2$, as follows immediately from (5.12). This enables us to move all of the $(1 - U_p p^{-s})^{-1}$ terms to the right past any $(1 - \chi(p_2)p_2^{k-1}V_{p_2}p_2^{-s})^{-1}$ term for $p_2 \neq p$. This gives us separate products with the V_p's and with the U_p's:

$$\prod_p \frac{1}{1 - \chi(p)p^{k-1}V_p p^{-s}} \prod_p \frac{1}{1 - U_p p^{-s}}.$$

We now expand each term in a geometric series and use the fact that $V_{mn} = V_m \circ V_n$ and $U_{mn} = U_m \circ U_n$. The result is (5.17). □

Proposition 39. *Under the conditions of Proposition* 37, *if* $T_m f(z) = \Sigma_{n=0}^{\infty} b_n q^n$, *then*

$$b_n = \sum_{d|\text{g.c.d.}(m,n)} \chi(d) d^{k-1} a_{mn/d^2}. \tag{5.19}$$

Proof. According to (5.18), we have

$$T_m \sum_{n=0}^{\infty} a_n q^n = \sum_{d|m} \chi(d) d^{k-1} V_d \circ U_{m/d} \sum_{n=0}^{\infty} a_n q^n$$
$$= \sum_{d|m} \chi(d) d^{k-1} \sum_{m/d|n} a_n q^{d^2 n/m}.$$

If we set $r = d^2 n/m$, the inner sum becomes $\Sigma\, a_{rm/d^2} q^r$ with the sum taken over all r divisible by d. Replacing r by n and gathering together coefficients of q^n, we obtain the expression (5.19) for the n-th coefficient. □

Most of the most important examples of modular forms turn out to be eigenvectors ("eigenforms") for the action of all of the T_m on the given space of modular forms. If $f \in M_k(N, \chi)$ is such an eigenform, then we can conclude a lot of information about its q-expansion coefficients.

Proposition 40. *Suppose that* $f(z) \in M_k(N, \chi)$ *is an eigenform for all of the operators* T_m *with eigenvalues* λ_m, $m = 1, 2, \ldots$: $T_m f = \lambda_m f$. *Let* a_m *be the* q*-expansion coefficients:* $f(z) = \Sigma_{m=0}^{\infty} a_m q^m$. *Then* $a_m = \lambda_m a_1$ *for* $m = 1, 2, \ldots$. *In addition,* $a_1 \neq 0$ *unless* $k = 0$ *and* f *is a constant function. Finally, if* $a_0 \neq 0$, *then* λ_m *is given by the formula*

$$\lambda_m = \sum_{d|m} \chi(d) d^{k-1}. \tag{5.20}$$

Proof. Using (5.19) with $n = 1$, we find that the coefficient of the first power of q in $T_m f$ is a_m. If $T_m f = \lambda_m f$, then this coefficient is also equal to $\lambda_m a_1$. This proves the first assertion. If we had $a_1 = 0$, then it would follow that all $a_m = 0$, and f would be a constant. Finally, suppose that $a_0 \neq 0$. If we compare the constant terms in $T_m f = \lambda_m f$ and use (5.19) with $n = 0$, we obtain: $\lambda_m a_0 = b_0 = \Sigma_{d|m} \chi(d) d^{k-1} a_0$. Dividing by a_0 gives the formula for λ_m. □

If f is an eigenform as in Proposition 40 (with $k \neq 0$), then we can multiply it by a suitable constant to get the coefficient of q equal to 1: $a_1 = 1$. Such an eigenform is called "normalized". In that case, Proposition 40 tells us that $a_m = \lambda_m$ is simply the eigenvalue of T_m. If we then apply the operator identity (5.11) to the eigenform f, we obtain identities for the q-expansion coefficients of f. Namely, applying both sides of (5.11) to a normalized eigenform $f \in M_k(N, \chi)$, we have:

$$\sum_{n=1}^{\infty} a_n n^{-s} = \prod_p (1 - a_p p^{-s} + \chi(p) p^{k-1-2s})^{-1}.$$

Examples. 1. Let $N=1$, and let $k \geq 4$ be an even integer. Suppose that $f=\Sigma a_n q^n \in M_k(\Gamma)$ is an eigenform for all of the T_m, and suppose that $a_0 \neq 0$ and f is normalized (i.e., $a_1=1$). Then Proposition 40 says that $f=a_0+\Sigma \sigma_{k-1}(n)q^n$. But recall the Eisenstein series $E_k=1-\frac{2k}{B_k}\Sigma \sigma_{k-1}(n)q^n$ (see (2.11)). Thus, f and $-\frac{B_k}{2k}E_k$ are two elements of $M_k(\Gamma)$ which differ by a constant. Since there are no nonzero constants in $M_k(\Gamma)$, they must be equal, i.e., $a_0=-B_k/2k$. Therefore, up to a constant factor, any weight k eigenform for Γ which is not a cusp form (i.e., $a_0 \neq 0$) must be E_k. Conversely, it is not hard to show that E_k is actually an eigenform for all of the T_m. This can be done using the q-expansion and (5.14) (it suffices to show that it is an eigenform for the T_p, since any T_m is a polynomial in the operators T_p for p prime). Another method is to use the original definition of T_m on modular points, applied to $G_k(L)=\Sigma_{0\neq l\in L}\, l^{-k}$.

2. Let $N=1$, $k=12$. Since $S_{12}(\Gamma)$ is one-dimensional, spanned by $(2\pi)^{-12}\Delta(z)=\Sigma_{n=1}^{\infty}\tau(n)q^n$ (see Propositions 9(d) and 15), and since the T_m preserve $S_k(\Gamma)$, it follows that $f=\Sigma\tau(n)q^n$ is an eigenform for T_m. It is normalized, since $\tau(1)=1$. Then Proposition 40 says that $T_m f=\tau(m)f$ for all m. Thus, the relation (5.11) applied to f gives

$$\sum_{n=1}^{\infty}\tau(n)n^{-s}=\prod_{\text{all }p}\frac{1}{1-\tau(p)p^{-s}+p^{11-2s}}. \tag{5.21}$$

This Euler product for the Mellin transform of $f=\Sigma\tau(n)q^n$ is equivalent to the sequence of identities (see Proposition 32):

$$\tau(mn)=\tau(m)\tau(n) \quad \text{for } m, n \text{ relatively prime};$$

$$\tau(p^l)=\tau(p^{l-1})\tau(p)-p^{11}\tau(p^{l-2}).$$

Ramanujan conjectured that in the denominator of (5.21) the quadratic $1-\tau(p)X+p^{11}X^2$ (where $X=p^{-s}$) has complex conjugate reciprocal roots α_p and $\bar{\alpha}_p$; equivalently, $1-\tau(p)X+p^{11}X^2=(1-\alpha_p X)(1-\bar{\alpha}_p X)$, i.e., $\tau(p)=\alpha_p+\bar{\alpha}_p$ with $|\alpha_p|=p^{11/2}$. From this it is easy to conclude that $|\tau(p)|\leq 2p^{11/2}$, and, more generally, $|\tau(n)|\leq\sigma_0(n)n^{11/2}$ (see the proof of Proposition 13 in §II.6). The Ramanujan conjecture and its generalization, the Ramanujan–Petersson conjecture for cusp forms which are eigenforms for the Hecke operators, were proved by Deligne as a consequence of the Weil conjectures (see [Katz 1976a]).

The Euler product (5.21) is reminiscent of the Hasse–Weil L-series for elliptic curves. For more information on such connections, see [Shimura 1971].

3. Let $N=4$, $\chi(n)=\left(\frac{-1}{n}\right)=(-1)^{(n-1)/2}$ for n odd. We saw that $M_1(N,\chi)$ is one-dimensional and is spanned by Θ^2. If we apply Proposition 40 to $\frac{1}{4}\Theta^2=\frac{1}{4}+q+\cdots+\lambda_m q^m+\cdots$, we find that $\lambda_m=\Sigma\left(\frac{-1}{d}\right)$, where the sum is over odd d dividing m. For example,

$$\lambda_p=1+\left(\frac{-1}{p}\right)=\begin{cases}0 & \text{if } p\equiv 3 \pmod 4;\\ 2 & \text{if } p\equiv 1 \pmod 4.\end{cases} \tag{5.22}$$

Since λ_m is one-fourth the number of ways m can be written as a sum of two squares, we have recovered the well-known fact that p can be so expressed in eight ways if $p \equiv 1 \pmod 4$ and in no way if $p \equiv 3 \pmod 4$. That fact is usually proved using factorization in the Gaussian integers $\mathbb{Z}[i]$. The Gaussian field $\mathbb{Q}(i)$ has discriminant -4, and corresponds to the character χ of $\mathbb{Z}$ having conductor 4: $\chi(n) = (\frac{-1}{n})$. Let 1 denote the trivial character of $(\mathbb{Z}/4\mathbb{Z})^*$, and let ρ denote the two-dimensional representation $(\begin{smallmatrix} 1 & 0 \\ 0 & \chi \end{smallmatrix})$ of $(\mathbb{Z}/4\mathbb{Z})^*$. Then we can write (5.22) in the form $\lambda_p = \operatorname{Tr} \rho(p)$. This turns out to be a very special case of a general fact: a normalized weight-one eigenform in $M_1(N, \chi)$ has its q-expansion coefficients determined by the trace of a certain two-dimensional representation of a certain galois group. For more information about this, see [Deligne and Serre 1974].

Another approach to Hecke operators. Let Γ_1 and Γ_2 be two subgroups of some group G.

Definition. Γ_1 and Γ_2 are said to be "commensurable" if their intersection has finite index in each group: $[\Gamma_1 : \Gamma_1 \cap \Gamma_2] < \infty$, $[\Gamma_2 : \Gamma_1 \cap \Gamma_2] < \infty$.

Basic Example. Let Γ' be a congruence subgroup of $\Gamma = SL_2(\mathbb{Z})$, and let $\alpha \in G = GL_2^+(\mathbb{Q})$. Then Γ' and $\alpha^{-1}\Gamma'\alpha$ are commensurable. To see this, suppose $\Gamma' \supset \Gamma(N)$. By Lemma 1 in the proof of Proposition 17, we have $\Gamma' \cap \alpha^{-1}\Gamma'\alpha \supset \Gamma(ND)$ and $\Gamma' \cap \alpha\Gamma'\alpha^{-1} \supset \Gamma(ND)$ for some D. Let $\Gamma'' = \Gamma' \cap \alpha^{-1}\Gamma'\alpha$. Then $\Gamma'' \supset \Gamma(ND)$, $\alpha\Gamma''\alpha^{-1} \supset \Gamma(ND)$. Thus, $[\Gamma' : \Gamma''] \le [\Gamma : \Gamma(ND)] < \infty$, and also $[\alpha^{-1}\Gamma'\alpha : \Gamma''] = [\Gamma' : \alpha\Gamma''\alpha^{-1}] \le [\Gamma : \Gamma(ND)] < \infty$.

Definition. If $\Gamma_1, \Gamma_2 \subset G$ and $\alpha \in G$, then the double coset $\Gamma_1 \alpha \Gamma_2$ is the set of all elements of G of the form $\gamma_1 \alpha \gamma_2$ with $\gamma_1 \in \Gamma_1$, $\gamma_2 \in \Gamma_2$. Notice that $\Gamma_1 \alpha \Gamma_2$ contains the right coset $\Gamma_1 \alpha$, and in general is a union of right cosets of the form $\Gamma_1 \alpha \gamma_2$.

Proposition 41. *Let $\Gamma' \subset G$ be any subgroup of a group, and let $\alpha \in G$ be any element such that Γ' and $\alpha^{-1}\Gamma'\alpha$ are commensurable. Let $\Gamma'' = \Gamma' \cap \alpha^{-1}\Gamma'\alpha$. Let $[\Gamma' : \Gamma''] = d$, and write $\Gamma' = \bigcup_{j=1}^d \Gamma''\gamma_j'$. Then $\Gamma'\alpha\Gamma' = \bigcup_{j=1}^d \Gamma'\alpha\gamma_j'$ is a disjoint union of d right cosets. Conversely, if $\Gamma'\alpha\Gamma' = \bigcup_{j=1}^d \Gamma'\alpha\gamma_j'$ is a disjoint union of d right cosets, then $\Gamma' = \bigcup_{j=1}^d \Gamma''\gamma_j'$.*

Proof. Given an element $\gamma_1\alpha\gamma_2$ with $\gamma_1, \gamma_2 \in \Gamma'$, we can write $\gamma_2 = \gamma''\gamma_j'$ with $\gamma'' \in \Gamma''$ for some j. Since $\gamma'' \in \alpha^{-1}\Gamma'\alpha$, we can write $\gamma'' = \alpha^{-1}\gamma'\alpha$, so that $\gamma_1\alpha\gamma_2 = \gamma_1\alpha(\alpha^{-1}\gamma'\alpha)\gamma_j' = (\gamma_1\gamma')\alpha\gamma_j' \in \Gamma'\alpha\gamma_j'$. We must show that the right cosets $\Gamma'\alpha\gamma_j'$ are distinct for different j. Suppose $\gamma_1\alpha\gamma_j' = \gamma_2\alpha\gamma_k'$. Then $\gamma_j'\gamma_k'^{-1} = \alpha^{-1}\gamma_1^{-1}\gamma_2\alpha \in \alpha^{-1}\Gamma'\alpha$. Since $\gamma_j'\gamma_k'^{-1} \in \Gamma'$, it follows that $\gamma_j'\gamma_k'^{-1} \in \Gamma''$, i.e., $\gamma_j' \in \Gamma''\gamma_k'$, and so $j = k$. The converse assertion is also routine, so we shall omit the details. □

We shall define Hecke operators in terms of double cosets for a large class of congruence subgroups of Γ, a class which includes $\Gamma_1(N)$, $\Gamma_0(N)$ and $\Gamma(N)$ as special cases. Then we shall show that in the case of $\Gamma_1(N)$ this definition coincides with our earlier definition of the operators T_n acting on $M_k(\Gamma_1(N))$.

Let S^+ be a nonzero additive subgroup of the integers, i.e., $S^+ = M\mathbb{Z}$ for some positive integer M. Let $S^\times$ be a subgroup of $(\mathbb{Z}/N\mathbb{Z})^*$. We shall also use $S^\times$ to denote the subset of $\mathbb{Z}$ whose image modulo N is in $S^\times$. (If $N = 1$, then we agree to take $S^\times = \mathbb{Z}$.) For example, if $S^\times = \{1\}$, then we also use $S^\times$ to denote $1 + N\mathbb{Z} \subset \mathbb{Z}$. Let n be any positive integer. We define

$$\Delta^n(N, S^\times, S^+) \underset{\text{def}}{=} \{\text{integer matrices } (\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}) |\ N|c, a \in S^\times, b \in S^+, \det (\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}) = n\}. \tag{5.23}$$

If $N = 1$ and $S^\times = S^+ = \mathbb{Z}$, then Δ^n is simply the set of all 2×2-matrices with determinant n. It is easy to check that

$$\Delta^m(N, S^\times, S^+) \cdot \Delta^n(N, S^\times, S^+) \subset \Delta^{mn}(N, S^\times, S^+), \tag{5.24}$$

and that $\Delta^1(N, S^\times, S^+)$ is a group. $\Delta^1(N, S^\times, S^+)$ is clearly a congruence subgroup of Γ, since it contains $\Gamma(N')$, where N' is the least common multiple of M and N (recall $S^+ = M\mathbb{Z}$). Here are our familiar examples:

$$\Gamma_1(N) = \Delta^1(N, \{1\}, \mathbb{Z}); \qquad \Gamma_0(N) = \Delta^1(N, (\mathbb{Z}/N\mathbb{Z})^*, \mathbb{Z});$$
$$\Gamma(N) = \Delta^1(N, \{1\}, N\mathbb{Z}).$$

Definition. Let Γ' be a congruence subgroup of Γ, and let $\alpha \in GL_2^+(\mathbb{Q})$. Let $\Gamma'' = \Gamma' \cap \alpha^{-1}\Gamma'\alpha$, and let $d = [\Gamma' : \Gamma'']$, $\Gamma' = \bigcup_{j=1}^{d} \Gamma''\gamma_j'$. Let $f(z)$ be a function on H which is invariant under $[\gamma]_k$ for $\gamma \in \Gamma'$. Then

$$f(z)|[\Gamma'\alpha\Gamma']_k \underset{\text{def}}{=} \sum_{j=1}^{d} f(z)|[\alpha\gamma_j']_k. \tag{5.25}$$

Proposition 42. *$f(z)|[\Gamma'\alpha\Gamma']_k$ does not change if α is replaced by any other representative α' of the same double coset: $\Gamma'\alpha'\Gamma' = \Gamma'\alpha\Gamma'$. Nor does it depend on the choice of representatives γ_j' of Γ' modulo Γ''. If $f \in M_k(\Gamma')$, then $f|[\Gamma'\alpha\Gamma']_k \in M_k(\Gamma')$.*

PROOF. We first prove the second assertion, that is, that (5.25) is unchanged if γ_j' is replaced by $\gamma_j''\gamma_j'$, where $\gamma_j'' \in \Gamma''$. Since $\Gamma'' \subset \alpha^{-1}\Gamma'\alpha$, we have $\gamma_j'' = \alpha^{-1}\tilde{\gamma}_j\alpha$ for some $\tilde{\gamma}_j \in \Gamma'$. Then $f|[\alpha\gamma_j''\gamma_j']_k = f|[\tilde{\gamma}_j\alpha\gamma_j']_k = f|[\alpha\gamma_j']_k$, because $f|[\tilde{\gamma}_j]_k = f$. Next, we observe that, by Proposition 41, the sum on the right in (5.25) can be written $\Sigma f(z)|[\alpha_j]_k$, where the α_j are any elements such that $\Gamma'\alpha\Gamma' = \bigcup \Gamma'\alpha_j$. It is then immediate that the definition depends only on the double coset $\Gamma'\alpha\Gamma'$ and not on the choice of representative α. Finally, suppose that $f \in M_k(\Gamma')$. If $\gamma \in \Gamma'$, then $(f|[\Gamma'\alpha\Gamma']_k)|[\gamma]_k = \Sigma f|[\alpha\gamma_j'\gamma]_k = f|[\Gamma'\alpha\Gamma']_k$, since right multiplication by γ just rearranges the right cosets $\Gamma'\alpha\gamma_j'$. If f

satisfies the cusp conditions, then so does $f|[\alpha\gamma_j']_k$ for each j, by Lemma 2 in the proof of Proposition 17. This completes the proof of Proposition 42. □

Definition. Let $\Gamma' = \Delta^1(N, S^\times, S^+)$, and let n be a positive integer. Let $f \in M_k(\Gamma')$. Then

$$T_n f \underset{\text{def}}{=} n^{(k/2)-1} \sum f|[\Gamma'\alpha\Gamma']_k, \tag{5.26}$$

where the sum is over all double cosets of Γ' in $\Delta^n(N, S^\times, S^+)$.

By Proposition 42, we have $T_n f \in M_k(\Gamma')$.

Equivalently, we can define

$$T_n f = n^{(k/2)-1} \sum f|[\alpha_j]_k, \tag{5.27}$$

where $\Gamma'\alpha_j$ runs through the right cosets of Γ' in $\Delta^n(N, S^\times, S^+)$.

Proposition 43. *In the case $\Gamma' = \Gamma_1(N)$, the definition* (5.26) *agrees with our earlier definition of the Hecke operators T_n.*

Proof. Let $\Delta^n = \Delta^n(N, \{1\}, \mathbb{Z})$. For each $a \in (\mathbb{Z}/N\mathbb{Z})^*$ we fix $\sigma_a \in \Gamma$ such that $\sigma_a \equiv \left(\begin{smallmatrix} 1/a & 0 \\ 0 & a \end{smallmatrix}\right) \bmod N$. □

Lemma.

$$\Delta^n = \bigcup_{\text{disjoint}} \Gamma_1(N)\sigma_a \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}, \tag{5.28}$$

where the disjoint union is taken over all positive a dividing n and prime to N, and for each such a we set $d = n/a$ and take $b = 0, 1, \ldots, d-1$.

Proof of Lemma. The terms on the right in (5.28) are clearly contained in Δ^n. Suppose the union were not disjoint. Then for some $a', b', d' = n/a'$ we would have $\gamma_1\sigma_a\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) = \gamma_2\sigma_{a'}\left(\begin{smallmatrix} a' & b' \\ 0 & d' \end{smallmatrix}\right)$, and so Γ would contain the matrix $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)\left(\begin{smallmatrix} a' & b' \\ 0 & d' \end{smallmatrix}\right)^{-1} = \left(\begin{smallmatrix} a/a' & * \\ 0 & d/d' \end{smallmatrix}\right)$; then $a' = a$, $d' = d$, and so $\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} 1 & j \\ 0 & 1 \end{smallmatrix}\right)\left(\begin{smallmatrix} a' & b' \\ 0 & d' \end{smallmatrix}\right)$ for some j, i.e., $b = b' + jd$. If $0 \le b, b' < d$, this means that $b = b'$.

To prove the lemma, it remains to show that any $\alpha = \left(\begin{smallmatrix} a' & b' \\ c' & d' \end{smallmatrix}\right) \in \Delta^n$ is in one of the terms on the right in (5.28). Choose g, h relatively prime so that $ga' + hc' = 0$, and complete the row g, h to a matrix $\gamma = \left(\begin{smallmatrix} e & f \\ g & h \end{smallmatrix}\right) \in \Gamma$. Then $\gamma\alpha$ has determinant n and is of the form $\left(\begin{smallmatrix} * & * \\ 0 & * \end{smallmatrix}\right)$. Replacing γ by $\pm\left(\begin{smallmatrix} 1 & j \\ 0 & 1 \end{smallmatrix}\right)\gamma$ if necessary, without loss of generality we may suppose that $\gamma\alpha = \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ with $a > 0$, $ad = n$, $0 \le b < d$. Then, considered modulo N, the equality $\alpha = \gamma^{-1}\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$ gives $\left(\begin{smallmatrix} 1 & * \\ 0 & n \end{smallmatrix}\right) \equiv \left(\begin{smallmatrix} 1/a & * \\ 0 & a \end{smallmatrix}\right)\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$, i.e., γ^{-1} is of the form $\gamma_1\sigma_a$ for some $\gamma_1 \in \Gamma_1(N)$. Thus, $\alpha = \gamma_1\sigma_a\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \in \Gamma_1(N)\sigma_a\left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)$, as desired. This completes the proof of the lemma. □

We now prove the proposition. Let T_n^{new} be the linear map on $M_k(\Gamma_1(N))$ defined by (5.26) (or (5.27)), and let T_n^{old} be the earlier definition. Since $M_k(\Gamma_1(N)) = \oplus_\chi M_k(N, \chi)$, it suffices to show that $T_n^{\text{new}} = T_n^{\text{old}}$ on $M_k(N, \chi)$.

Let $f \in M_k(N, \chi)$. By the lemma, we have

$$T_n^{\text{new}} f = n^{(k/2)-1} \sum f | [\sigma_a \begin{pmatrix} a & b \\ 0 & d \end{pmatrix}]_k.$$

Since $f \in M_k(N, \chi)$ we have $f | [\sigma_a]_k = \chi(a) f$. Also, for each $a > 0$ and d with $ad = n$ we have

$$\sum_{b=0}^{d-1} f(z) | [\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}]_k = \sum_{b=0}^{d-1} n^{k/2} d^{-k} f\left(\frac{az+b}{d}\right)$$
$$= n^{k/2} d^{-k} V_a \circ d U_d f(z),$$

by (5.13). Thus,

$$T_n^{\text{new}} f = n^{(k/2)-1} \sum_{a|n} \chi(a) n^{k/2} d^{1-k} V_a \circ U_d f.$$

(The sum is over $a > 0$ dividing n and prime to N; but the term $\chi(a)$ ensures that the terms with g.c.d.$(a, N) > 1$ will drop out.) That is,

$$T_n^{\text{new}} = \sum_{a|n} \chi(a) a^{k-1} V_a \circ U_{n/a} = T_n^{\text{old}},$$

by (5.18). This proves Proposition 43. □

In many situations, the definition of T_n in terms of double cosets is more convenient than the definition in terms of modular points. For example, we shall use this definition below to show that T_n is a Hermitian operator with respect to the Petersson scalar product. Moreover, the double coset approach can be generalized to situations where no good interpretation in terms of modular points is known. We shall see an example of this in the next chapter, where we shall define Hecke operators on forms of half-integral weight. For a more detailed treatment of the double coset approach in a more general context, see [Shimura 1971].

The Petersson scalar product. Suppose we have an integral over some region in the upper half-plane and make the change of variable $z \mapsto \alpha z = \frac{az+b}{cz+d}$, where $\alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2^+(\mathbb{Q})$. If we have a term $dxdy/y^2$ in the integrand, this does not change under such a change of variable. To see this, we compute:

$$\frac{d\alpha z}{dz} = \frac{\det \alpha}{(cz+d)^2}; \qquad \frac{\operatorname{Im} \alpha z}{\operatorname{Im} z} = \frac{\det \alpha}{|cz+d|^2}. \tag{5.29}$$

In general, if we make a differentiable change of complex variable $z_1 = u(z)$, then the area element $dxdy$ near z is multiplied by $|u'(z)|^2$ (see Fig. III.5). Thus, interpreting the first equality in (5.29) in terms of the real variables $x = \operatorname{Re} z$ and $y = \operatorname{Im} z$, we see that an element of area near z is expanded by a factor $|d\alpha z/dz|^2 = (\det \alpha)^2/|cz+d|^4$. Thus, $dxdy/y^2$ is invariant under the change of variables, by (5.29).

Proposition 44. *Let $\Gamma' \subset \Gamma$ be a congruence subgroup, let $F' \subset H$ be any fundamental domain for Γ', and define*

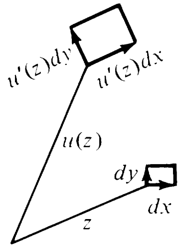


Figure III.5

$$\mu(\Gamma') \underset{\text{def}}{=} \int_{F'} \frac{dxdy}{y^2}. \tag{5.30}$$

Then

(a) *The integral* (5.30) *converges and is independent of the choice of* F'.
(b) $[\bar{\Gamma} : \bar{\Gamma}'] = \mu(\Gamma')/\mu(\Gamma)$.
(c) *If* $\alpha \in GL_2^+(\mathbb{Q})$ *and* $\alpha^{-1}\Gamma'\alpha \subset \Gamma$, *then* $[\bar{\Gamma} : \bar{\Gamma}'] = [\bar{\Gamma} : \alpha^{-1}\bar{\Gamma}'\alpha]$.

Proof. First let $[\bar{\Gamma} : \bar{\Gamma}'] = d$, let $\bar{\Gamma} = \bigcup_{j=1}^{d} \alpha_j \bar{\Gamma}'$, and take $F' = \bigcup \alpha_j^{-1} F$. Notice that the integral for $\mu(\Gamma)$ converges, because

$$\int_F \frac{dxdy}{y^2} < \int_{-1/2}^{1/2} \int_{\sqrt{3}/2}^{\infty} y^{-2} dy dx = \frac{2}{\sqrt{3}}.$$

If for each j we make the change of variables $z \mapsto \alpha_j z$, we find that $\int_{\alpha_j^{-1}F} dxdy/y^2 = \int_F dxdy/y^2$; hence, for our choice of F' we find that the integral in (5.30) converges to $d\mu(\Gamma)$. Suppose we chose a different fundamental domain F_1 for Γ'. Then we divide F_1 into regions R for which there exists $\alpha \in \Gamma'$ with $\alpha R \subset F'$, and again we use the invariance of $dxdy/y^2$ under $z \mapsto \alpha z$ to show that the integral over $R \subset F_1$ and the integral over the corresponding region αR in F' are equal. Finally, to show part (c), we note that $\alpha^{-1}F'$ is a fundamental domain for $\alpha^{-1}\Gamma'\alpha$. Since

$$\int_{\alpha^{-1}F'} \frac{dxdy}{y^2} = \int_{F'} \frac{dxdy}{y^2},$$

it follows that $\mu(\alpha^{-1}\Gamma'\alpha) = \mu(\Gamma')$, and so part (c) follows from part (b). □

Now suppose that $f(z)$ and $g(z)$ are two functions in $M_k(\Gamma')$. We consider the function $f(z)\overline{g(z)}y^k$, where the bar denotes complex conjugation and $y = \operatorname{Im} z$. If we replace the variable z by αz for $\alpha \in GL_2^+(\mathbb{Q})$, we obtain: $f(\alpha z)\overline{g(\alpha z)}y^k(\det \alpha/|cz + d|^2)^k$, by (5.29). But this is just $(f(z)|[\alpha]_k)$ $\overline{(g(z)|[\alpha]_k)}y^k$. Thus, the effect of the change of variable is to replace f by $f|[\alpha]_k$ and g by $g|[\alpha]_k$.

Definition. Let $\Gamma' \subset \Gamma$ be a congruence subgroup, let F' be a fundamental domain for Γ', and let $f, g \in M_k(\Gamma')$, with at least one of the two functions f,

g a cusp form. Then we define

$$\langle f, g\rangle \underset{\text{def}}{=} \frac{1}{[\bar{\Gamma}:\bar{\Gamma}']}\int_{F'} f(z)\overline{g(z)}y^k\frac{dxdy}{y^2}. \tag{5.31}$$

It is immediate from this definition that $\langle f, g\rangle$ is linear in f and antilinear in g (i.e., $\langle f, cg\rangle = \bar{c}\langle f, g\rangle$), it is antisymmetric (i.e., $\langle g, f\rangle = \overline{\langle f, g\rangle}$), and also $\langle f, f\rangle > 0$ for $f \neq 0$. These are the properties one needs in order to have a *hermitian* scalar product. We shall return to this later.

Proposition 45. *The integral in* (5.31) *is absolutely convergent, and does not depend on the choice of* F'. *If* Γ'' *is another congruence subgroup such that* $f, g \in M_k(\Gamma'')$, *then the definition of* $\langle f, g\rangle$ *is independent of whether* f, g *are considered in* $M_k(\Gamma')$ *or in* $M_k(\Gamma'')$.

Remark. The term $1/[\bar{\Gamma}:\bar{\Gamma}']$ in (5.31) is needed in order to have the second assertion in the proposition. The requirement that f or g be a cusp form is needed to get convergence of the integral, as we shall see.

PROOF. First take $F' = \bigcup \alpha_j^{-1}F$, where $\bar{\Gamma} = \bigcup \alpha_j\bar{\Gamma}'$. In each region $\alpha_j^{-1}F$ make the change of variables which replaces z by $\alpha_j^{-1}z$, thereby transforming the integral into

$$\sum_j \int_F f(z)|[\alpha_j^{-1}]_k\overline{g(z)|[\alpha_j^{-1}]_k}y^k\frac{dxdy}{y^2}.$$

Since $f, g \in M_k(\Gamma')$, we can write $f|[\alpha_j^{-1}]_k = \Sigma_{n=0}^{\infty} a_n q_N^n$, $g|[\alpha_j^{-1}]_k = \Sigma_{n=0}^{\infty} b_n q_N^n$, with either $a_n = 0$ or $b_n = 0$, since f or g is a cusp form. Because $|q_N| = |e^{2\pi iz/N}| = e^{-2\pi y/N}$ and $y \geq \frac{\sqrt{3}}{2}$ in F, and because a_n and b_n have only polynomial growth (see Problem 19 in §3), it is not hard to see that the integral is absolutely convergent. Next, if F_1 is another fundamental domain, we proceed as in the proof of Proposition 44, comparing $\int_R f(z)\overline{g(z)}y^{k-2}dxdy$ with $\int_{\alpha R} f(z)\overline{g(z)}y^{k-2}dxdy$, where R is a subregion of F_1 and αR, with $\alpha \in \Gamma'$, is the corresponding subregion of F'. Making the change of variable $z \mapsto \alpha z$ and using the fact that $f|[\alpha]_k = f$, $g|[\alpha]_k = g$, we find that the two integrals are equal.

Finally, suppose $f, g \in M_k(\Gamma'')$. First suppose that $\Gamma'' \subset \Gamma'$. Writing $\bar{\Gamma}' = \bigcup_l \delta_l\bar{\Gamma}''$, $F'' = \bigcup \delta_l^{-1}F'$, we have

$$\frac{1}{[\bar{\Gamma}:\bar{\Gamma}'']}\int_{F''} f(z)\overline{g(z)}y^k\frac{dxdy}{y^2}$$
$$= \frac{1}{[\bar{\Gamma}:\bar{\Gamma}']}\cdot\frac{1}{[\bar{\Gamma}':\bar{\Gamma}'']}\sum_l \int_{F'} f(z)|[\delta_l^{-1}]_k\overline{g(z)|[\delta_l^{-1}]_k}y^k\frac{dxdy}{y^2}.$$

But all of the summands on the right are equal, since $f|[\delta_l^{-1}]_k = f$, $g|[\delta_l^{-1}]_k = g$; and there are $[\bar{\Gamma}':\bar{\Gamma}'']$ values of l. We thus obtain the right side of (5.31).

If $\Gamma'' \not\subset \Gamma'$, we simply set $\Gamma''' = \Gamma'' \cap \Gamma'$, and show that (5.31) and (5.31) with Γ'' in place of Γ' are each equal to (5.31) with Γ''' in place of Γ'. This completes the proof of Proposition 45. □

Proposition 46. *Let $f, g \in M_k(\Gamma')$ with f or g a cusp form. Let $\alpha \in GL_2^+(\mathbb{Q})$. Then*

$$\langle f, g\rangle = \langle f|[\alpha]_k, g|[\alpha]_k\rangle. \tag{5.32}$$

Proof. Let $\Gamma'' = \Gamma' \cap \alpha\Gamma'\alpha^{-1}$. Then f, $g \in M_k(\Gamma'')$, and $f|[\alpha]_k$, $g|[\alpha]_k \in M_k(\alpha^{-1}\Gamma''\alpha)$ by Proposition 17(a). Here $\alpha^{-1}\Gamma''\alpha = \alpha^{-1}\Gamma'\alpha \cap \Gamma'$. Let F'' be a fundamental domain for Γ'', and take $\alpha^{-1}F''$ as a fundamental domain for $\alpha^{-1}\Gamma''\alpha$. Then the right side of (5.32) is

$$\frac{1}{[\bar{\Gamma}: \alpha^{-1}\bar{\Gamma}''\alpha]}\int_{\alpha^{-1}F''} f(z)|[\alpha]_k\overline{g(z)|[\alpha]_k}y^k\frac{dxdy}{y^2}$$
$$= \frac{1}{[\bar{\Gamma}: \alpha^{-1}\bar{\Gamma}''\alpha]}\int_{F''} f(z)\overline{g(z)}y^k\frac{dxdy}{y^2}.$$

Since $[\bar{\Gamma}: \alpha^{-1}\bar{\Gamma}''\alpha] = [\bar{\Gamma}: \bar{\Gamma}'']$ by Proposition 44(c), we obtain the left side of (5.32). □

Now for $\alpha \in GL_2^+(\mathbb{Q})$ we note that α can be multiplied by a positive scalar without affecting $[\alpha]_k$. So without loss of generality we shall assume in what follows that $\alpha = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right)$ has integer entries. Let $D = ad - bc = \det \alpha$, and set $\alpha' = D\alpha^{-1} = \left(\begin{smallmatrix} d & -b \\ -c & a\end{smallmatrix}\right)$. Then $[\alpha^{-1}]_k = [\alpha']_k$.

Proposition 47. *With f, g, α as in Proposition 46,*

$$\langle f|[\alpha]_k, g\rangle = \langle f, g|[\alpha']_k\rangle. \tag{5.33}$$

In addition, $\langle f|[\alpha]_k, g\rangle$ depends only on the double coset of α modulo Γ'.

Proof. If in (5.32) we replace g by $g|[\alpha^{-1}]_k$ and replace Γ' by $\Gamma' \cap \alpha\Gamma'\alpha^{-1}$, then Proposition 46 gives (5.33). Now suppose we replace α by $\gamma_1\alpha\gamma_2$ in $\langle f|[\alpha]_k, g\rangle$. We obtain $\langle f|[\gamma_1\alpha\gamma_2]_k, g\rangle = \langle f|[\gamma_1\alpha]_k, g|[\gamma_2^{-1}]_k\rangle = \langle f|[\alpha]_k, g\rangle$, because f is invariant under $[\gamma_1]_k$ and g is invariant under $[\gamma_2^{-1}]_k$, since $\gamma_1, \gamma_2^{-1} \in \Gamma'$. □

Proposition 48. *Let $\Gamma' = \Gamma_1(N)$, $\Delta^n = \Delta^n(N, \{1\}, \mathbb{Z})$ (see (5.23)). Let f, $g \in M_k(\Gamma')$ with f or g a cusp form. For each $d \in (\mathbb{Z}/N\mathbb{Z})^*$, fix some $\sigma_d \in \Gamma$ such that $\sigma_d \equiv \left(\begin{smallmatrix} 1/d & 0 \\ 0 & d\end{smallmatrix}\right)$ mod N. Let n be a positive integer prime to N. Then $\langle T_n f, g\rangle = \langle f|[\sigma_n]_k, T_n g\rangle$. In particular, if $f \in M_k(N, \chi)$, then $\langle T_n f, g\rangle = \chi(n)\langle f, T_n g\rangle$ for n prime to N.*

Proof. If $\alpha = \left(\begin{smallmatrix} a & b \\ c & d\end{smallmatrix}\right) \in \Delta^n$ then $\alpha \equiv \left(\begin{smallmatrix} 1 & b \\ 0 & n\end{smallmatrix}\right)$ mod N, $\alpha' \equiv \left(\begin{smallmatrix} n & -b \\ 0 & 1\end{smallmatrix}\right)$ mod N, and $\sigma_n\alpha' \equiv \left(\begin{smallmatrix} 1 & -b/n \\ 0 & n\end{smallmatrix}\right)$ mod N. Thus, $\sigma_n\alpha' \in \Delta^n$. By definition,

$$T_n f = n^{(k/2)-1}\sum f|[\Gamma'\alpha\Gamma']_k,$$

where the sum is over the distinct double cosets of $\Gamma' = \Delta^1$ in Δ^n. Let d_α denote the number of right cosets in $\Gamma'\alpha\Gamma'$. Then $d_\alpha = [\Gamma': \Gamma' \cap \alpha^{-1}\Gamma'\alpha]$ by Proposition 41. By (5.26), (5.25) and the second assertion in Proposition 47, we have

$$\langle T_n f, g\rangle = n^{(k/2)-1} \sum d_\alpha \langle f|[\alpha]_k, g\rangle,$$

where the sum takes one α from each double coset. The map $\Gamma'\alpha\Gamma' \mapsto \Gamma'(\sigma_n\alpha')\Gamma'$ permutes the double cosets. Namely, if we had $\sigma_n\alpha_1'$ and $\sigma_n\alpha_2'$ in the same double coset $\gamma_1\sigma_n\alpha_1'\gamma_2 = \sigma_n\alpha_2'$, then taking inverses and multiplying by n would give

$$\gamma_2^{-1}\alpha_1\sigma_n^{-1}\gamma_1^{-1} = \alpha_2\sigma_n^{-1},$$

i.e., $\alpha_1 = \gamma_2\alpha_2(\sigma_n{}^{-1}\gamma_1\sigma_n)$. Since $\sigma_n^{-1}\Gamma_1(N)\sigma_n = \Gamma_1(N)$, we have $\alpha_1 \in \Gamma'\alpha_2\Gamma'$, and so α_1 and α_2 are in the same double coset. Next, we observe that $d_{\sigma_n\alpha'} = d_\alpha$, because

$$\Gamma' \cap (\sigma_n\alpha')^{-1}\Gamma'\sigma_n\alpha' = \Gamma' \cap \alpha\sigma_n^{-1}\Gamma'\sigma_n\alpha^{-1} = \Gamma' \cap \alpha\Gamma'\alpha^{-1}$$
$$= \alpha(\alpha^{-1}\Gamma'\alpha \cap \Gamma')\alpha^{-1},$$

and so

$$d_\alpha = [\Gamma' : \Gamma' \cap \alpha^{-1}\Gamma'\alpha] = [\Gamma' : \Gamma' \cap (\sigma_n\alpha')^{-1}\Gamma'\sigma_n\alpha'] = d_{\sigma_n\alpha'},$$

by Proposition 44(c). Thus,

$$\langle T_n f, g\rangle = n^{(k/2)-1} \sum d_{\sigma_n\alpha'} \langle f|[\sigma_n\alpha']_k, g\rangle = n^{(k/2)-1} \sum d_\alpha \langle f|[\sigma_n]_k, g|[\alpha]_k\rangle$$

by Proposition 47. This equals $\langle f|[\sigma_n]_k, T_n g\rangle$, as desired. Finally, if $f \in M_k(N, \chi)$, then $f|[\sigma_n] = \chi(n)f$, so we obtain the final equality. □

A basis of eigenforms. We first remark that for any congruence subgroup Γ' and any integer k, the $\mathbb{C}$-vector space $M_k(\Gamma')$ is finite dimensional. One way to show this would be to take a fundamental domain F' for Γ' of the form $F' = \bigcup \alpha_j^{-1}F$ and integrate $f'(z)/f(z)$ around the boundary for nonzero $f \in M_k(\Gamma')$, as in the proof of Proposition 8 of §III.2, thereby obtaining a bound on the total number of zeros of f in a fundamental domain. But if $M_k(\Gamma')$ were infinite dimensional, by taking suitable linear combinations one could obtain nonzero $f \in M_k(\Gamma')$ which vanish at any given finite set of points in F'. Alternately, one could prove finite dimensionality, and even obtain formulas for the dimension, using the Riemann–Roch theorem (see [Shimura 1971, §2.6]).

The Petersson scalar product (5.31) gives a hermitian scalar product on the finite dimensional $\mathbb{C}$-vector space $S_k(\Gamma')$. That is, $\langle f, g\rangle$ is linear in f and antilinear in g, it is antisymmetric, and $\langle f, f\rangle > 0$ for $f \neq 0$, as we remarked when we gave the definition (5.31).

Proposition 50. *Let n be a positive integer prime to N, and let χ be a Dirichlet character modulo N. Let c_n be either square root of $\bar{\chi}(n)$. Then the operator $c_n T_n$ on $S_k(N, \chi)$ is hermitian, i.e., $\langle c_n T_n f, g\rangle = \langle f, c_n T_n g\rangle$.*

PROOF. $\langle c_n T_n f, g\rangle = c_n\langle T_n f, g\rangle = c_n\chi(n)\langle f, T_n g\rangle = c_n\bar{c}_n^2\langle f, T_n g\rangle = \bar{c}_n\langle f, T_n g\rangle = \langle f, c_n T_n g\rangle$, as claimed. □

We saw before that eigenforms for the T_n have nice properties: the coefficients can be expressed in terms of the eigenvalues for the T_p, and the corresponding Dirichlet series have Euler products. Because of Proposition 50, it is possible to find a basis of such forms.

Proposition 51. *There exists a basis of the $\mathbb{C}$-vector space $S_k(N, \chi)$ whose elements are eigenforms for all of the T_n for which* g.c.d.$(n, N) = 1$.

PROOF. For any fixed T_n with g.c.d.$(n, N) = 1$ and any subspace $S \subset S_k(N, \chi)$ which is preserved by T_n, there exists a basis of S consisting of eigenforms of T_n. To see this, we apply the following basic fact from linear algebra (see [Lang 1984, §XIV.12]) to the hermitian operator $c_n T_n$: Given any hermitian operator T on a finite dimensional $\mathbb{C}$-vector space S, there exists a basis of S consisting of eigenvectors for T. We further note that any eigenspace for T_n is preserved by all $T_{n'}$, as follows from the fact that T_n and $T_{n'}$ commute: if $T_n f = \lambda_n f$, then $T_n(T_{n'} f) = T_{n'} T_n f = \lambda_n T_{n'} f$. (This remark does not require n or n' to be prime to N.) Thus, to prove the proposition, we list the T_n for n prime to N (actually, it suffices to work with the T_p for primes $p \nmid N$, since any eigenform for them is an eigenform for the T_n with n prime to N). We write $S_k(N, \chi)$ as a direct sum of eigenspaces S for the first T_n in the list. Then we write each S as a direct sum of eigenspaces for the next T_n; then we write each one of those spaces (which is an eigenspace for the first two T_n's) as a sum of eigenspaces for the third T_n; and so on. Because $S_k(N, \chi)$ is finite dimensional, after finitely many steps this process must stop giving us any new smaller spaces. At that point $S_k(N, \chi)$ is expressed as a direct sum of subspaces on each of which the T_n for n prime to N act as a scalar. Any basis consisting of forms in these subspaces will satisfy the requirements of the proposition. □

Proposition 51 does not quite give us what we want, because of the restriction that n be prime to N. In order to have an Euler product (as, for example, in (5.21)), we would also want our basis forms to be eigenforms for T_p for $p \mid N$. One way we could ensure this is if we found that the eigenspaces for all of the T_n with n prime to N are each one-dimensional, because we know that the T_p for $p \mid N$ commute with the T_n and so preserve each eigenspace. Such an assertion is called "multiplicity one", i.e., each set of eigenvalues for the T_n with n prime to N corresponds to only one eigenform in the basis that was constructed in Proposition 51. Multiplicity one does not hold in general; however, it does hold if we restrict our attention to forms which do not come from lower level. We now explain what this means.

If $d_1 d_2 = N$ and $f \in M_k(\Gamma_1(d_1))$, then we also have $f \in M_k(\Gamma_1(N))$ and also $g(z) \underset{\text{def}}{=} f(d_2 z) \in M_k(\Gamma_1(N))$ (see Proposition 17). The subspace of $S_k(\Gamma_1(N))$ spanned by the forms obtained in these two ways from forms $f \in S_k(\Gamma_1(d))$ for proper divisors $d \mid N$ is called the space of "old forms". It is not hard to show that all Hecke operators preserve this space. The orthogonal comple-

ment to the space of old forms with respect to the Petersson scalar product is called the space of "new forms". In other words, a form $f \in S_k(\Gamma_1(N))$ is a new form if and only if $\langle f, g \rangle = 0$ and $\langle f, g | [\begin{smallmatrix} d_2 & 0 \\ 0 & 1 \end{smallmatrix}]_k \rangle = 0$ for every $g \in S_k(\Gamma_1(d_1))$, where $d_1 d_2 = N$, $1 < d_1 < N$. It can then be shown that the space of new forms has multiplicity one; thus, Proposition 51 holds for the space of new forms without the condition that n be prime to N. For details, see Chapter VIII of [Lang 1976].

Another question not answered by Proposition 51 is the existence of a basis of eigenforms for all of $M_k(\Gamma_1(N))$. Here we cannot consider $\langle \ , \ \rangle$ as a scalar product on $M_k(\Gamma_1(N))$, because $\langle f, g \rangle$ makes sense only if f or g is a cusp form. Instead, one can use the explicit construction of Eisenstein series in §III.3 above, directly constructing eigenforms. Then $M_k(\Gamma_1(N))$ can be written as the orthogonal direct sum of $S_k(\Gamma_1(N))$ and a space spanned by Eisenstein series which are eigenforms. (In fact, an intrinsic definition of an Eisenstein series, which generalizes to many other situations, is: a modular form which is orthogonal to all cusp forms.) The dimension of the space of Eisenstein series turns out to be the number r of regular cusps of $\Gamma_1(N)$. Roughly speaking, in order for a modular form to be a cusp form it must satisfy r conditions (vanishing of the constant term in the q_N-expansion), one at each regular cusp; and so $S_k(\Gamma_1(N))$ has codimension r in $M_k(\Gamma_1(N))$. For more information on Eisenstein series, see, for example, [Gunning 1962].

Problems

1. Prove Proposition 31 in cases (iii)–(iv), i.e., $\Gamma' = \Gamma_0(N)$, $\Gamma(N)$.

2. (a) Let $\alpha_N = (\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix})$. We saw in Problem 15 of §III.3 that $[\alpha_N]_k$ preserves $M_k(\Gamma_0(N))$. Prove that the Hecke operator T_n commutes with α_N for n prime to N.
   (b) Let $F = \Sigma_{n\,\mathrm{odd}}\, \sigma_1(n) q^n$, $\Theta^4 = \Sigma_n a_n q^n \in M_2(\Gamma_0(4))$ (where a_n equals the number of ways n can be written as a sum of four squares). Write the matrix of T_2 in the basis Θ^4, F, and find a basis of normalized eigenforms for T_2.
   (c) Show that T_2 does *not* commute with α_4 (see Problem 15(c) in §III.3).
   (d) Suppose n is odd. Since T_n commutes with α_4 and T_2, it preserves each eigenspace in part (b) and each eigenspace in Problem 15(c) of §III.3. Show that the operator T_n on the two-dimensional space $M_2(\Gamma_0(4))$ is simply multiplication by $\sigma_1(n)$.
   (e) Derive the following famous formula (see, for example, Chapter 20 of [Hardy and Wright 1960]) for the number of ways n can be written as a sum of four squares:

$$a_n = \begin{cases} 8\sigma_1(n) & \text{for } n \text{ odd}; \\ 24\sigma_1(n_0) & \text{for } n = 2^r n_0 \text{ even, } 2 \nmid n_0. \end{cases}$$

3. If $f \in M_k(N, \chi)$, show that $f(\infty) = 0$ implies $T_n f(\infty) = 0$, but that $f(s) = 0$ does *not* necessarily imply $T_n f(s) = 0$ for other cusps s (give an example).

4. (a) Show that if $f \in M_k(N, \chi)$ and $g(z) = f(Mz)$, then $g \in M_k(MN, \chi')$, where χ' is defined by $\chi'(n) = \chi(n \bmod N)$ for $n \in (\mathbb{Z}/MN\mathbb{Z})^*$. Let T_n be the Hecke operator

considered on $M_k(N, \chi)$, and let T_n' be the Hecke operator considered on $M_k(MN, \chi')$. Suppose n is prime to M. Show that for $f \in M_k(N, \chi) \subset M_k(MN, \chi')$ and $g(z) = f(Mz) \in M_k(MN, \chi')$ we have:

$$T_n' f = T_n f; \qquad T_n' g(z) = (T_n f)(Mz).$$

(b) Let $f(z) = (\eta(z)\eta(2z))^8$, $g(z) = f(2z) \in S_8(\Gamma_0(4))$ (see Problem 18(d) of §III.3). Show that for odd n, T_n acts on the two-dimensional space $S_8(\Gamma_0(4))$ by multiplication by a scalar.

(c) Find the matrix of T_2 acting on $S_8(\Gamma_0(4))$ in the basis f, g; then find the basis of normalized eigenforms for all of the T_n.

5. Let $f, g \in M_k(N, \chi)$ with f or g a cusp form. Let $p|N$. Note that $T_p = U_p$. Show that $\langle T_p f, g\rangle = p^k \langle f, V_p g\rangle$.

6. Let $f_1 = (2\pi)^{-24}\Delta^2$, $f_2 = (2\pi)^{-12}\Delta E_6^2$, where $(2\pi)^{-12}\Delta = q\Pi_n(1 - q^n)^{24}$ and $E_6 = 1 - 504\,\Sigma\,\sigma_5(n)q^n$. We know that $S_{24}(\Gamma)$ is two-dimensional and is spanned by f_1 and f_2 (see §III.2).
   (a) Give a simple reason why f_1 is *not* an eigenform for the Hecke operators.
   (b) With the help of a calculator, find the matrix of the Hecke operator T_2 in the basis f_1, f_2.
   (c) Express in terms of f_1 and f_2 the basis for $S_{24}(\Gamma)$ consisting of normalized eigenforms for all the T_n.
   (d) Express in terms of f_1 and f_2 the cusp form whose n-th q-expansion coefficient is the trace of T_n on $S_{24}(\Gamma)$.
   (e) Find Tr T_3 from part (d), and also directly by computing matrix entries.

In this problem one encounters fairly large numbers; for example, the coefficients in part (c) are in $\mathbb{Q}(\sqrt{144169})$. While Proposition 51 guarantees the existence of a basis of normalized eigenforms, it does not guarantee that it will always be easy to find it computationally.

7. Let (L, t) be a modular point for $\Gamma_1(N)$. Let $\alpha \in \Delta^n(N, \{1\}, \mathbb{Z})$, and let $\gamma \in \Gamma_1(N)$. For each α and γ, consider the lattice $L' = \frac{1}{n}\alpha\gamma L$. Show that L' is a lattice which contains L with index n, that L' depends only on the right coset of $\Gamma_1(N)$ in $\Delta^n(N, \{1\}, \mathbb{Z})$ which contains $\alpha\gamma$, and that t has order N in $\mathbb{C}/L'$. Show that the L' in (5.2) are in one-to-one correspondence with the right cosets of $\Gamma_1(N)$ in $\Delta^n(N, \{1\}, \mathbb{Z})$. Use this to give another proof of Proposition 43 by comparing (5.9) and (5.27).

8. Let $f \in M_k(\Gamma_0(p))$.
   (a) Show that $\gamma_0 = 1$ and $\gamma_j = \left(\begin{smallmatrix} 0 & -1 \\ 1 & j \end{smallmatrix}\right)$, $0 \le j < p$, are right coset representatives for Γ modulo $\Gamma_0(p)$.
   (b) Let $\operatorname{Tr}(f)$ be defined by: $\operatorname{Tr}(f) = \Sigma_{j=0}^{p} f|[\gamma_j]_k$. Show that $\operatorname{Tr}(f) \in M_k(\Gamma)$.
   (c) If f happens to be in $M_k(\Gamma)$, then show that $\operatorname{Tr}(f) = (p+1)f$, and $\operatorname{Tr}(f|[\left(\begin{smallmatrix} 0 & -1 \\ p & 0 \end{smallmatrix}\right)]_k) = p^{1-k/2}T_p f$.

CHAPTER IV

Modular Forms of Half Integer Weight

Let k be a positive odd integer, and let $\lambda = (k - 1)/2$. In this chapter we shall look at modular forms of weight $k/2 = \lambda + 1/2$, which is not an integer but rather half way between two integers. Roughly speaking, such a modular form f should satisfy $f((az + b)/(cz + d)) = (cz + d)^{\lambda+1/2}f(z)$ for $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ in $\Gamma = SL_2(\mathbb{Z})$ or some congruence subgroup $\Gamma' \subset \Gamma$. However, such a simple-minded functional equation leads to inconsistencies (see below), basically because of the possible choice of two branches for the square root. A subtler definition is needed in order to handle the square root properly. One must introduce a quadratic character, corresponding to some quadratic extension of $\mathbb{Q}$. Roughly speaking, because of this required "twist" by a quadratic character, the resulting forms turn out to have interesting relationships to the arithmetic of quadratic fields (such as L-series and class numbers). Moreover, recall that the Hasse–Weil L-series for our family of elliptic curves $E_n: y^2 = x^3 - n^2x$ in the congruent number problem involved "twists" by quadratic characters as n varies (see Chapter II). It turns out that the critical values $L(E_n, 1)$ for this family of L-series are closely related to certain modular forms of half-integral weight.

One classical reason why modular forms were studied is their use in investigating the number of ways of representing an integer by a quadratic form: $m = \sum_{j,l=1}^{k} A_{jl}n_jn_l = [n]^tA[n]$, where $A = [A_{jl}]$ is a given symmetric matrix, $[n]$ is a column vector and $[n]^t$ the corresponding row vector. For example, the number of ways m can be represented as a sum of k squares is equal to the coefficient a_m in the q-expansion of Θ^k:

$$\Theta^k = \prod_{j=1}^{k} \sum_{n_j=-\infty}^{\infty} q^{n_j^2} = \sum_{n_1,\cdots,n_k=-\infty}^{\infty} q^{\Sigma n_j^2} = \sum a_m q^m.$$

In Chapter III we saw that for k even $\Theta^k \in M_{k/2}(4, \chi_{-1}^{k/2})$, where χ_{-1} is the

character modulo 4 defined by $\chi_{-1}(n) = (\frac{-1}{n}) = (-1)^{(n-1)/2}$. More generally, for k even one can use $\Sigma_n q^{[n]^t A[n]}$ to construct modular forms of weight $k/2$. The properties of modular forms are then useful in studying the number of representations $m = [n]^t A[n]$. For example, in Problem 2 of §III.5 we used the action of the Hecke operators on $M_2(\Gamma_0(4))$ to derive a simple formula for the number of ways an integer can be written as a sum of four squares. For more information about the connection between quadratic forms in k variables and modular forms of weight $k/2$, see, for example, [Gunning 1962] and [Ogg 1969].

It is natural to ask whether a similar theory exists for quadratic forms in an odd number of variables. This would be a theory of modular forms of weight $k/2$ where k is an *odd* integer. One would want Θ^k for k odd to be an example of such a form. Early investigators of representability of an integer as a sum of an odd number of squares—Eisenstein, and later G. H. Hardy—understood the desirability of such a theory of modular forms of half-integral weight. But it was only recently—starting with Shimura's 1973 *Annals* paper—that major advances have been made toward a systematic theory of such forms which rivals in elegance and beauty the much older theory of forms of integral weight.

In this chapter we shall first present the basic definitions and elementary properties of forms of half-integral weight, largely following [Shimura 1973a, 1973b], but with slightly different notation. We shall discuss examples in some detail. But when we come to the fundamental theorems of Shimura and Waldspurger, we shall not give the proofs, which go beyond the scope of an introductory text. Rather, we shall motivate those theorems using the examples. Finally, we shall conclude by returning to the congruent number problem and Tunnell's theorem, which we used in Chapter I to motivate our study of elliptic curves.

§1. Definitions and examples

We shall always take the branch of the square root having argument in $(-\pi/2, \pi/2]$. Thus, $\sqrt{z}$ is a holomorphic function on the complex plane with the negative real axis $(-\infty, 0]$ removed. It takes positive reals to positive reals, complex numbers in the upper half-plane to the first quadrant, and complex numbers in the lower half-plane to the fourth quadrant. For any integer k, we define $z^{k/2}$ to mean $(\sqrt{z})^k$.

Whenever we have a transformation rule such as $f(\gamma z) = (cz + d)^k f(z)$ (where $\gamma = (\begin{smallmatrix} a & b \\ c & d \end{smallmatrix})$, $\gamma z = (az + b)/(cz + d)$), we call the term $(cz + d)^k$ the "automorphy factor". It depends on γ and on z. That is, an automorphy factor $J(\gamma, z)$ for a nonzero function f has the property that $f(\gamma z) = J(\gamma, z) f(z)$, for $z \in H$ and γ in some matrix group. Because

$$\frac{f(\alpha\beta z)}{f(z)} = \frac{f(\alpha\beta z)}{f(\beta z)} \cdot \frac{f(\beta z)}{f(z)},$$

it follows that any automorphy factor $J(\gamma, z)$ must satisfy

$$J(\alpha\beta, z) = J(\alpha, \beta z) \cdot J(\beta, z). \tag{1.1}$$

For example, $J(\gamma, z) = cz + d$ for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma = SL_2(\mathbb{Z})$ satisfies (1.1). If $-1 = \left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$ is in our matrix group, then $J(\gamma, z)$ must also satisfy: $J(-1, z) = 1$. Another example of an automorphy factor is $J(\gamma, z) = j(\gamma, z) = \left(\frac{c}{d}\right)\varepsilon_d^{-1}\sqrt{cz + d}$, $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(4)$, which is the automorphy factor for $\Theta(z) = \sum_{n=-\infty}^{\infty} q^{n^2}$, $q = e^{2\pi i z}$ (see §III.4).

Suppose that, in defining modular forms of weight $k/2$ for k an odd integer, we try the most obvious thing: we look for functions f satisfying $f(\gamma z) = (cz + d)^{k/2} f(z)$. However, $J(\gamma, z) = (cz + d)^{k/2}$ cannot be an automorphy factor of a nonzero function for *any* congruence subgroup $\Gamma' \subset \Gamma$ and any odd integer k. To see this, suppose $\Gamma' \supset \Gamma(N)$, $N > 2$. Let $\alpha = \left(\begin{smallmatrix} N+1 & N \\ -N & 1-N \end{smallmatrix}\right)$, $\beta = \left(\begin{smallmatrix} 1 & 0 \\ -N & 1 \end{smallmatrix}\right)$. Then $\alpha, \beta \in \Gamma'$, and (1.1) would require the k-th power of the following equality of holomorphic functions on H:

$$\sqrt{(N^2 - 2N)z + 1 - N} = \sqrt{-Nz/(-Nz + 1) + 1 - N} \cdot \sqrt{-Nz + 1}. \tag{1.2}$$

Clearly, the square of (1.2) holds; thus (1.2) itself holds up to a sign. Since both expressions in the radicals on the right are in the lower half-plane, the right side is the product of two complex numbers in the fourth quadrant; but the left side is in the first quadrant, since $(N^2 - 2N)z + 1 - N \in H$. Hence, (1.2) is off by a factor of -1, and (1.1), which is the k-th power of (1.2), is also off by -1 if k is odd. Thus, we cannot have $J(\gamma, z) = (cz + d)^{k/2}$.

The natural way out of this difficulty is to force (1.1) to hold by simply defining the automorphy factor $J(\gamma, z)$ to be the k-th power of

$$j(\gamma, z) \underset{\text{def}}{=} \Theta(\gamma z)/\Theta(z) = \left(\frac{c}{d}\right) \varepsilon_d^{-1} \sqrt{cz + d} \quad \text{for} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(4). \tag{1.3}$$

That is, for a congruence subgroup $\Gamma' \subset \Gamma_0(4)$, one defines a modular form of weight $k/2$ to be *a holomorphic function on H which transforms like the k-th power of $\Theta(z)$ under any fractional linear transformation in* Γ' (and which is holomorphic at the cusps, in a sense to be explained below).

Recall that $\left(\frac{c}{d}\right)$ is the Legendre symbol, defined for a positive odd prime d as the usual quadratic residue symbol, then for all positive odd d by multiplicativity, and finally for negative odd d as $\left(\frac{c}{|d|}\right)$ if $c > 0$ and $-\left(\frac{c}{|d|}\right)$ if $c < 0$. (Also, $\left(\frac{0}{\pm 1}\right) = 1$.) Further recall that we define $\varepsilon_d = 1$ if $d \equiv 1 \pmod 4$ and $\varepsilon_d = i$ if $d \equiv -1 \pmod 4$. Thus, $\varepsilon_d^2 = \chi_{-1}(d) = (-1)^{(d-1)/2}$ (note that this holds for negative as well as positive d).

Thus, if we define $f(z)|[\gamma]_{k/2} = j(\gamma, z)^{-k} f(\gamma z)$ for $\gamma \in \Gamma_0(4)$ and k odd, we shall require a modular form of weight $k/2$ for Γ' to be fixed by $[\gamma]_{k/2}$ for $\gamma \in \Gamma'$. As in the case of integer weight, we shall want to define $[\alpha]_{k/2}$ for arbitrary matrices $\alpha \in GL_2^+(\mathbb{Q})$ with rational entries and positive determinant.

But since the automorphy factor $j(\gamma, z)$ is defined only for $\gamma \in \Gamma_0(4)$, we have no preferred branch of the square root for arbitrary $\alpha \in GL_2^+(\mathbb{Q})$. This circumstance requires us to work with a bigger group than $GL_2^+(\mathbb{Q})$, a group G that contains two "copies" of each $\alpha \in GL_2^+(\mathbb{Q})$, one corresponding to each branch of the square root of $cz + d$. The example of the automorphy factor $j(\gamma, z)$, which is a square root of $(\frac{-1}{d})(cz + d)$, shows that we also should allow square roots of $-(cz + d)$. Thus, we shall actually define G to be a *four*-sheeted covering of $GL_2^+(\mathbb{Q})$. We now give the precise definition of the group G.

Let $T \subset \mathbb{C}$ denote the group of fourth roots of unity: $T = \{\pm 1, \pm i\}$. We now define G to be the set of all ordered pairs $(\alpha, \phi(z))$, where $\alpha = (\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}) \in GL_2^+(\mathbb{Q})$ and $\phi(z)$ is a holomorphic function on H such that

$$\phi(z)^2 = t\frac{cz + d}{\sqrt{\det \alpha}}$$

for some $t \in T^2 = \{\pm 1\}$. That is, for each $\alpha \in GL_2^+(\mathbb{Q})$ and for each fixed $t = \pm 1$, there are two possible elements $(\alpha, \pm\phi(z)) \in G$. We define the product of two elements of G as follows:

$$(\alpha, \phi(z))(\beta, \psi(z)) = (\alpha\beta, \phi(\beta z)\psi(z)). \tag{1.4}$$

Remark. We could define G in the same way but with T defined to be the group of *all* roots of unity or even (as in [Shimura 1973a]) the group of all complex numbers of absolute value 1. However, for our purposes we only need fourth roots of unity.

Proposition 1. *G is a group under the operation* (1.4).

PROOF. We first check closure, i.e., that the right side of (1.4) belongs to G. If $\alpha = (\begin{smallmatrix} a & b \\ c & d \end{smallmatrix})$, $\beta = (\begin{smallmatrix} e & f \\ g & h \end{smallmatrix})$, we have

$$\begin{aligned}(\phi(\beta z)\psi(z))^2 &= t_1(c\beta z + d)t_2(gz + h)/\sqrt{\det \alpha \det \beta} \\ &= t_1 t_2(c(ez + f) + d(gz + h))/\sqrt{\det \alpha\beta} \\ &= t_1 t_2((ce + dg)z + cf + dh)/\sqrt{\det \alpha\beta},\end{aligned}$$

which shows that $(\alpha\beta, \phi(\beta z)\psi(z)) \in G$. Associativity is immediate. It is also obvious that $((\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}), 1) \in G$ serves as the identity element. Finally, to find an inverse of $(\alpha, \phi(z))$, where $\alpha = (\begin{smallmatrix} a & b \\ c & d \end{smallmatrix})$, we write $D = \det \alpha$, $\alpha' = D\alpha^{-1} = (\begin{smallmatrix} d & -b \\ -c & a \end{smallmatrix})$, and look for $(\alpha^{-1}, \psi(z))$ such that: $\phi(\alpha^{-1}z)\psi(z) = 1$. That is, we set $\psi(z) = 1/\phi(\alpha^{-1}z)$, and then must check that $(\alpha^{-1}, \psi(z)) \in G$. But

$$\begin{aligned}\psi(z)^2 &= 1\Big/\left(t\left(c\frac{dz - b}{-cz + a} + d\right)\Big/\sqrt{D}\right) = \frac{1}{t}\sqrt{D}(-cz + a)/(ad - bc) \\ &= \frac{1}{t}\left(-\frac{c}{D}z + \frac{a}{D}\right)\Big/\sqrt{\det \alpha^{-1}},\end{aligned}$$

which is of the required form. This completes the proof. □

We have a homomorphism

$$P\colon G \to GL_2^+(\mathbb{Q})$$

which simply projects onto the first part of the pair: $(\alpha, \phi(z)) \mapsto \alpha$. The kernel of P is the set of all $(1, \phi(z)) \in G$. Since $\phi(z)^2 = t$ for $(1, \phi(z)) \in G$, it follows that $\phi(z)$ must be a constant function equal to a fourth root of unity. Thus, if we map T to G by $t \mapsto (1, t)$, we have an exact sequence

$$1 \to T \to G \xrightarrow{P} GL_2^+(\mathbb{Q}) \to 1;$$

in other words, the kernal of P is the image of T under $t \mapsto (1, t)$.

Similarly, if $\alpha = a\left(\begin{smallmatrix}1 & j\\ 0 & 1\end{smallmatrix}\right)$ for $a \in \mathbb{Q}$, the inverse image of α under P is the set of all pairs (α, t) for $t \in T$, since in that case we again find that $\phi(z)^2$ must be ± 1.

We let $G^1 = P^{-1}(\Gamma)$ be the set of all pairs $(\alpha, \phi(z)) \in G$ such that $\alpha \in \Gamma = SL_2(\mathbb{Z})$. G^1 is clearly a subgroup. If $\xi = (\alpha, \phi(z)) \in G$, we shall sometimes use the notation ξz to mean the same as αz for $z \in H$.

For $\xi = (\alpha, \phi(z)) \in G$ and any integer k, we define an operator $[\xi]_{k/2}$ on functions f on the upper half-plane H by the rule

$$f(z)|[\xi]_{k/2} \underset{\text{def}}{=} f(\alpha z)\phi(z)^{-k}. \tag{1.5}$$

This gives an action of the group G on the space of such functions, i.e., we have $(f(z)|[\xi_1]_{k/2})|[\xi_2]_{k/2} = f(z)|[\xi_1\xi_2]_{k/2}$, because of (1.4).

Now let Γ' be a subgroup of $\Gamma_0(4)$. Then $j(\gamma, z)$ is defined for elements $\gamma \in \Gamma'$ (see (1.3)). We define

$$\tilde{\Gamma}' \underset{\text{def}}{=} \{(\gamma, j(\gamma, z)) | \gamma \in \Gamma'\}. \tag{1.6}$$

Clearly $\tilde{\Gamma}'$ is a subgroup of G (because of (4.13) in §III.4) which is isomorphic to Γ' under the projection P. Let L denote the map from $\Gamma_0(4)$ to G which takes γ to

$$\tilde{\gamma} \underset{\text{def}}{=} (\gamma, j(\gamma, z)) \in G.$$

Then P and L are mutually inverse isomorphisms from $\tilde{\Gamma}_0(4)$ to $\Gamma_0(4)$. We call L a "lifting" or "section" of the projection P. In our notation we are using tildes to denote this lifting from $\Gamma_0(4)$ to G:

$$L\colon \gamma \mapsto \tilde{\gamma} = (\gamma, j(\gamma, z));$$

$$P\colon (\gamma, j(\gamma, z)) \mapsto \gamma.$$

Suppose that $f(z)$ is a meromorphic function on H which is invariant under $[\tilde{\gamma}]_{k/2}$ for $\gamma \in \Gamma'$, where Γ' is a subgroup of finite index in $\Gamma_0(4)$. We now describe what it means for $f(z)$ to be meromorphic, holomorphic, or vanish at a cusp $s \in \mathbb{Q} \cup \{\infty\}$. We first treat the cusp ∞. Since Γ' has finite index in $\Gamma_0(4)$ and hence in Γ, its intersection with

$$\Gamma_\infty = \left\{\pm\begin{pmatrix}1 & j\\ 0 & 1\end{pmatrix}\right\}_{j \in \mathbb{Z}}$$

must be of the form $\{\pm\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right)^j\}$ if $-1\in\Gamma'$ and either $\{\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right)^j\}$ or $\{(-\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right))^j\}$ if $-1\notin\Gamma'$. (We always take $h>0$.) Since $[\widetilde{-1}]_{k/2}$ is the identity, and $j(\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right),z)=1$, in all cases f is invariant under $z\mapsto z+h$, and so has an expansion in powers of $q_h=e^{2\pi iz/h}$. As in the case of integer weight, we say that f is meromorphic at ∞ if only finitely many negative powers of q_h occur, we say that f is holomorphic at ∞ if no negative power of q_h occurs, and we define $f(\infty)$ to be the constant term when f is holomorphic at ∞.

Now suppose that $s\in\mathbb{Q}\cup\{\infty\}$. Let $s=\alpha\infty$, $\alpha\in\Gamma$, and let $\xi=(\alpha,\phi(z))$ be any element of G^1 which projects to α: $P(\xi)=\alpha$. Define $\tilde{\Gamma}'_s=\{\tilde{\gamma}\in\tilde{\Gamma}'|\gamma s=s\}$, and let $G^1_\infty=\{\eta\in G^1|\eta\infty=\infty\}$. We have

$$G^1_\infty=\left\{\left(\pm\begin{pmatrix}1&j\\0&1\end{pmatrix},t\right)\middle|\, j\in\mathbb{Z},\ t\in T\right\},$$

because $\Gamma_\infty=\{\pm\left(\begin{smallmatrix}1&j\\0&1\end{smallmatrix}\right)\}$, and the $\phi(z)$ for $\left(\begin{smallmatrix}1&j\\0&1\end{smallmatrix}\right)$ must be a constant function t, as noted above. Now $\xi^{-1}\tilde{\Gamma}'_s\xi$ is contained in G^1 and fixes ∞, i.e., $\xi^{-1}\tilde{\Gamma}'_s\xi\subset G^1_\infty$. Moreover, P gives an isomorphism from $\xi^{-1}\tilde{\Gamma}'_s\xi$ to $\alpha^{-1}\Gamma'_s\alpha\subset\Gamma_\infty$. Since Γ' has finite index in Γ, it follows that $\alpha^{-1}\Gamma'_s\alpha$ is of one of the forms $\{\pm\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right)^j\}$, $\{\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right)^j\}$, or $\{(-\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right))^j\}$. Thus, for some $t\in T$ we have

$$\pm\xi^{-1}\tilde{\Gamma}'_s\xi=\left\{\left(\pm\begin{pmatrix}1&h\\0&1\end{pmatrix},t\right)^j\right\}_{j\in\mathbb{Z}}.$$

Given s and ξ, $h>0$ is determined by requiring $\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right)$ to be a generator of $\alpha^{-1}\Gamma'_s\alpha$ modulo ±1; then t is determined, since P is one-to-one on $\xi^{-1}\tilde{\Gamma}'_s\xi$, i.e., we can find t by applying the lifting map L to $\pm\alpha\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right)\alpha^{-1}\in\Gamma'_s$.

Proposition 2. *The element $(\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right),t)\in G^1_\infty$ depends only on the Γ'-equivalence class of the cusp s.*

Proof. First suppose that ξ is replaced by another element $\xi_1=(\alpha_1,\phi_1(z))\in G^1$ such that $s=\alpha_1\infty$. Then $\xi^{-1}\xi_1\in G^1$ fixes ∞, and so it is of the form $(\pm\left(\begin{smallmatrix}1&l\\0&1\end{smallmatrix}\right),t_1)$. Then

$$\begin{aligned}\pm\xi_1^{-1}\tilde{\Gamma}'_s\xi_1&=(\xi^{-1}\xi_1)^{-1}(\pm\xi^{-1}\tilde{\Gamma}'_s\xi)(\xi^{-1}\xi_1)\\&=\left(\begin{pmatrix}1&l\\0&1\end{pmatrix},t_1\right)\left\{\left(\pm\begin{pmatrix}1&h\\0&1\end{pmatrix},t\right)^j\right\}\left(\begin{pmatrix}1&-l\\0&1\end{pmatrix},t_1^{-1}\right).\end{aligned}$$

But $(\left(\begin{smallmatrix}1&l\\0&1\end{smallmatrix}\right),t)$ commutes with $(\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right),t)$, so conjugating by $\xi^{-1}\xi_1$ does not affect $(\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right),t)$.

Now suppose that $s_1=\gamma s=(\tilde{\gamma}\xi)\infty$, where $\gamma\in\Gamma'$, $\tilde{\gamma}=(\gamma,j(\gamma,z))\in\tilde{\Gamma}'$. Note that $\Gamma'_{s_1}=\gamma\Gamma'_s\gamma^{-1}$, and so $\tilde{\Gamma}'_{s_1}=\tilde{\gamma}\tilde{\Gamma}'_s\tilde{\gamma}^{-1}$. Thus,

$$\pm(\tilde{\gamma}\xi)^{-1}\tilde{\Gamma}'_{s_1}\tilde{\gamma}\xi=\pm\xi^{-1}\tilde{\gamma}^{-1}(\tilde{\gamma}\tilde{\Gamma}'_s\tilde{\gamma}^{-1})\tilde{\gamma}\xi=\pm\xi^{-1}\tilde{\Gamma}'_s\xi,$$

and we again obtain the same $(\left(\begin{smallmatrix}1&h\\0&1\end{smallmatrix}\right),t)$. This completes the proof. □

We are now ready to define meromorphicity/holomorphicity/vanishing at the cusp s of Γ'. Suppose f is invariant under $[\tilde{\gamma}]_{k/2}$ for $\tilde{\gamma}\in\tilde{\Gamma}'$. Let $s=\xi\infty$, and set $g=f|[\xi]_{k/2}$. Then for any element $\pm\xi^{-1}\tilde{\gamma}\xi\in\pm\xi^{-1}\tilde{\Gamma}'\xi$ we have

$$g|[\pm\xi^{-1}\tilde{\gamma}\xi]_{k/2}=f|[\tilde{\gamma}\xi]_{k/2}=f|[\xi]_{k/2}=g.$$

That is, g is invariant under $[((\begin{smallmatrix}1&h\\0&1\end{smallmatrix}),t)]_{k/2}$:

$$g(z)=g(z)\left|\left[\left(\begin{pmatrix}1&h\\0&1\end{pmatrix},t\right)\right]_{k/2}=t^{-k}g(z+h).\right.$$

Write $t^k=e^{2\pi ir}$, where $r=0,\frac{1}{4},\frac{1}{2}$, or $\frac{3}{4}$. Then $e^{-2\pi irz/h}g(z)$ is invariant under $z\mapsto z+h$, and so we obtain a Fourier series expansion $e^{-2\pi irz/h}g(z)=\Sigma\,a_nq_h^n$, i.e.,

$$g(z)=\sum_n a_ne^{2\pi iz(n+r)/h}.$$

We say that f is meromorphic at s if $a_n=0$ for all but finitely many $n<0$, and that f is holomorphic at s if $a_n=0$ for all $n<0$. If f is holomorphic at s, we define $f(s)\underset{\text{def}}{=}\lim_{z\to i\infty}g(z)$. Automatically $f(s)=0$ if $r\neq 0$, since in that case the first term is $a_0e^{2\pi izr/h}$. If $r=0$, then $f(s)=a_0$. It is easy to see that these definitions depend only on the Γ'-equivalence class of s, i.e., $g=f|[\xi]_{k/2}$ may be replaced by $g_1=f|[\tilde{\gamma}\xi_1]_{k/2}$, where $\tilde{\gamma}\in\tilde{\Gamma}'$ and $\xi_1\infty=s$ (see the proof of Proposition 2, and also the proof of Proposition 16 in §III.3).

It should be noted, however, that there is some indeterminacy in the value $f(s)$ of a modular form at a cusp. Namely, if we replace $\xi\in G^1$ by $((\begin{smallmatrix}1&0\\0&1\end{smallmatrix}),t_1)\xi$, with $t_1\in T$, then the effect is to multiply $g=f|[\xi]_{k/2}$ by t_1^{-k}. If $k/2$ is a half-integer, this may alter the value by a power of i; if $k/2$ is an odd integer, then $f(s)$ may be defined only up to ±1 (compare with the proof of Proposition 16); and if $k/2$ is an even integer, then $f(s)$ is well defined in all cases.

Given a cusp s of Γ' and an integer k, we say that s is *k-irregular* if $r\neq 0$, and we say that s is *k-regular* if $r=0$, i.e., if $t^k=1$. Thus, if f is holomorphic at the cusps, it automatically vanishes at all k-irregular cusps.

Note that whether a given cusp s of Γ' is k-regular or k-irregular depends only on k modulo 4. That is, if $t=\pm i$, then the cusp is k-irregular unless $k/2$ is an even integer; if $t=-1$, then it is k-irregular unless $k/2$ is an integer; and if $t=1$, then it is always k-regular. In the case when $k/2$ is an odd integer, the terminology agrees with the definition of regular and irregular cusps in Problem 2 of §III.3.

Definitions. Let k be any integer, and let $\Gamma'\subset\Gamma_0(4)$ be a subgroup of finite index. Let $f(z)$ be a meromorphic function on the upper half-plane H which is invariant under $[\tilde{\gamma}]_{k/2}$ for all $\tilde{\gamma}\in\tilde{\Gamma}'$. We say that $f(z)$ is a *modular function of weight* $k/2$ for $\tilde{\Gamma}'$ if f is meromorphic at every cusp of Γ'. We say that such an $f(z)$ is a modular form and write $f\in M_{k/2}(\tilde{\Gamma}')$, if it is holomorphic on H and at every cusp. We say that a modular form f is a cusp form and write $f\in S_{k/2}(\tilde{\Gamma}')$, if it vanishes at every cusp.

Now let N be a positive multiple of 4, so that $\Gamma_0(N) \subset \Gamma_0(4)$. Let χ be a character of $(\mathbb{Z}/N\mathbb{Z})^*$. We let $M_{k/2}(\tilde{\Gamma}_0(N), \chi)$ denote the subspace of $M_{k/2}(\tilde{\Gamma}_1(N))$ consisting of f such that for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$

$$f|[\tilde{\gamma}]_{k/2} = \chi(d)f. \tag{1.7}$$

We define $S_{k/2}(\tilde{\Gamma}_0(N), \chi) = S_{q/2}(\tilde{\Gamma}_1(N)) \cap M_{k/2}(\tilde{\Gamma}_0(N), \chi)$. The exact same argument as in the integer weight case shows that

$$M_{k/2}(\tilde{\Gamma}_1(N)) = \oplus_\chi M_{k/2}(\tilde{\Gamma}_0(N), \chi).$$

Furthermore, it follows immediately from the definitions that if χ_1 and χ_2 are Dirichlet characters modulo N, then

$$f_i \in M_{k_i/2}(\tilde{\Gamma}_0(N), \chi_i) \qquad (i = 1, 2) \quad \text{implies} \quad f_1 f_2 \in M_{(k_1+k_2)/2}(\tilde{\Gamma}_0(N), \chi_1\chi_2). \tag{1.8}$$

Notice that for any $k \in \mathbb{Z}$, we have $M_{k/2}(\tilde{\Gamma}_0(N), \chi) = 0$ if χ is an odd character, as we see by substituting -1 for γ in (1.7): $f|[(\left(\begin{smallmatrix} -1 & 0 \\ 0 & -1 \end{smallmatrix}\right), 1)]_{k/2} = f = \chi(-1)f$. For example, since the trivial character $\chi = 1$ is the only even character of $(\mathbb{Z}/4\mathbb{Z})^*$, this means that

$$M_{k/2}(\tilde{\Gamma}_1(4)) = M_{k/2}(\tilde{\Gamma}_0(4), 1) = M_{k/2}(\tilde{\Gamma}_0(4)).$$

If $4|N$, recall that χ_{-1} denotes the character modulo N defined by $\chi_{-1}(n) = (-1)^{(n-1)/2}$ for $n \in (\mathbb{Z}/N\mathbb{Z})^*$.

Notice that in our definitions k is any integer, not necessarily odd. For k even, let us compare the above definitions of $M_{k/2}(\tilde{\Gamma}')$, $S_{k/2}(\tilde{\Gamma}')$, $M_{k/2}(\tilde{\Gamma}_0(N), \chi)$, $S_{k/2}(\tilde{\Gamma}_0(N), \chi)$ with the definitions of $M_{k/2}(\Gamma')$, $S_{k/2}(\Gamma')$, $M_{k/2}(N, \chi)$, $S_{k/2}(N, \chi)$ in Chapter III. First, for any $\xi = (\alpha, \phi(z)) \in G$, $\alpha = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$, $\phi(z)^2 = t(cz+d)/\sqrt{\det \alpha}$, note that $f(z)|[\xi]_{k/2} = \phi(z)^{-k} f(\alpha z) = t^{-k/2} f(z)|[\alpha]_{k/2}$, so that $[\xi]_{k/2}$ for $k/2 \in \mathbb{Z}$ differs only by a root of unity from the operator $[\alpha]_{k/2}$ defined in the last chapter. It immediately follows that for $k/2 \in \mathbb{Z}$ the cusp condition defined in the last chapter is equivalent to the cusp condition defined above.

There is a slight difference, however, between the condition that f be invariant under $[\gamma]_{k/2}$ for $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma'$ and the condition that f be invariant under $[\tilde{\gamma}]_{k/2}$. Namely, $f|[\tilde{\gamma}]_{k/2} = j(\gamma, z)^{-k} f(\gamma z) = ((\frac{-1}{d})(cz+d))^{-k/2} f(\gamma z)) = (\frac{-1}{d})^{k/2} f|[\gamma]_{k/2}$. Thus, if $k/2$ is odd, then $[\tilde{\gamma}]_{k/2}$ differs from $[\gamma]_{k/2}$ by $\chi_{-1}(d)$. From this discussion we conclude

Proposition 3. *Let* $4|N$, $k/2 \in \mathbb{Z}$. *Then*

$$M_{k/2}(\tilde{\Gamma}_0(N), \chi) = M_{k/2}(N, \chi_{-1}^{k/2}\chi), \qquad S_{k/2}(\tilde{\Gamma}_0(N), \chi) = S_{k/2}(N, \chi_{-1}^{k/2}\chi).$$

It is now not hard to describe the structure of $M_{k/2}(\tilde{\Gamma}_0(4))$. If we have a polynomial $P(X_1, \ldots, X_m) \in \mathbb{C}[X_1, \ldots, X_m]$ and assign "weights" w_i to X_i, then we say that a monomial $\Pi X_i^{n_i}$ has weight $w = \Sigma n_i w_i$, and we say that a polynomial P has pure weight w if every monomial which occurs in P with nonzero coefficient has weight w.

Proposition 4. *Let* $\Theta(z) = \Sigma_{n=-\infty}^{\infty} q^{n^2}$, $F(z) = \Sigma_{n>0\text{ odd}} \sigma_1(n)q^n$, $q = e^{2\pi iz}$. *Assign weight* $1/2$ *to* Θ *and weight* 2 *to* F. *Then* $M_{k/2}(\tilde{\Gamma}_0(4))$ *is the space of all polynomials in* $\mathbb{C}[\Theta, F]$ *having pure weight* $k/2$.

Proof. In Chapter III we found that for $k/2 \in \mathbb{Z}$, the space $M_{k/2}(N, \chi_{-1}^{k/2})$ consists of polynomials in Θ and F having pure weight $k/2$ (see Problems 17–18 of §III.3). This gives Proposition 4 when $k/2 \in \mathbb{Z}$. Next, by the definition of $[\tilde{\gamma}]_{k/2}$, we have $\Theta|[\tilde{\gamma}]_{1/2} = \Theta$ for $\gamma \in \Gamma_0(4)$. Θ is holomorphic at the cusps, because we checked holomorphicity of Θ^2 at the cusps, and if Θ had a singularity at s, so would Θ^2. Thus, $\Theta \in M_{1/2}(\tilde{\Gamma}_0(4))$. It now follows by (1.8) that any polynomial in Θ and F of pure weight $k/2$ is in $M_{k/2}(\tilde{\Gamma}_0(4))$. Conversely, suppose that $f \in M_{k/2}(\tilde{\Gamma}_0(4))$, where k is odd. Then $f\Theta$ is in $M_{(k+1)/2}(\tilde{\Gamma}_0(4))$, and so can be written in the form $f\Theta = aF^{(k+1)/4} + \Theta^2 P(\Theta, F)$, where $a \in \mathbb{C}$ is a constant which equals 0 if $4 \nmid k+1$ and $P(\Theta, F)$ has pure weight $(k-1)/2$. Then $f - \Theta P(\Theta, F) = aF^{(k+1)/4}/\Theta \in M_{k/2}(\tilde{\Gamma}_0(4))$. But because Θ vanishes at the cusp $-\frac{1}{2}$, while F does not, this form satisfies the holomorphicity condition at $-\frac{1}{2}$ only if $a = 0$. Thus, $f = \Theta P(\Theta, F)$, and the proof is complete. □

Corollary. Dim $M_{k/2}(\tilde{\Gamma}_0(4)) = 1 + [k/4]$.

Problems

1. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(4)$, and let $\rho = (\left(\begin{smallmatrix} m & 0 \\ 0 & 1 \end{smallmatrix}\right), m^{-1/4})$. Check that $\rho \in G$.
   (a) Compute $\rho\tilde{\gamma}\rho^{-1}$.
   (b) If $\gamma \in \Gamma_0(4m)$, then $\gamma_1 = \left(\begin{smallmatrix} m & 0 \\ 0 & 1 \end{smallmatrix}\right)\gamma\left(\begin{smallmatrix} m & 0 \\ 0 & 1 \end{smallmatrix}\right)^{-1}$ is in $\Gamma_0(4)$. Compare $\rho\tilde{\gamma}\rho^{-1}$ with $\tilde{\gamma}_1$. Show that they are always the same if m is a perfect square, but that otherwise they differ by a sign for certain $\gamma \in \Gamma_0(4m)$.
2. Let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(4)$, and let $\rho = (\left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right), N^{1/4}\sqrt{z})$. Check that $\rho \in G$.
   (a) Compute $\rho\tilde{\gamma}\rho^{-1}$.
   (b) Suppose that $4|N$ and $\gamma \in \Gamma_0(N)$. Then $\gamma_1 = \left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right)\gamma\left(\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}\right)^{-1} \in \Gamma_0(N)$. Compare $\rho\tilde{\gamma}\rho^{-1}$ with $\tilde{\gamma}_1$.
3. Find h and t for each cusp of $\Gamma_0(4)$.
4. Let $\Gamma' \subset \Gamma_0(4)$ be a subgroup of finite index. Let $k/2 \in \mathbb{Z}$. Show that if $\Gamma' \subset \Gamma_1(4)$, then $M_{k/2}(\tilde{\Gamma}') = M_{k/2}(\Gamma')$. Otherwise, let χ be the unique nontrivial character of $\Gamma'/\Gamma' \cap \Gamma_1(4)$; show that $M_{k/2}(\tilde{\Gamma}') = M_{k/2}(\Gamma', \chi^{k/2})$ in the notation of §III.3 (see the discussion following Proposition 27).
5. (a) Describe $S_{k/2}(\tilde{\Gamma}_0(4))$, and find its dimension.
   (b) Show that for $k \geq 5$, the codimension of $S_{k/2}(\tilde{\Gamma}_0(4))$ in $M_{k/2}(\tilde{\Gamma}_0(4))$ is equal to the number of k-regular cusps.
   (c) Find an element $\Sigma\, a_n q^n$ in $S_{13/2}(\tilde{\Gamma}_0(4))$ such that $a_1 = 1$ and $a_2 = 0$ (there is only one).
6. Let $4|N$, and define χ_N by $\chi_N(d) = \left(\frac{N}{d}\right)$ if g.c.d.$(N, d) = 1$, $\chi_N(d) = 0$ if g.c.d.$(N, d) > 1$.
   (a) Show that χ_N is a Dirichlet character modulo N.

(b) Let ρ be as in Problem 2 above, let χ be an arbitrary Dirichlet character modulo N, and suppose that $f \in M_{k/2}(\tilde{\Gamma}_0(N), \chi)$. Show that $f|[\rho]_{k/2} \in M_{k/2}(\tilde{\Gamma}_0(N), \bar{\chi}\chi_N^k)$.

7. Find the value of $\Theta \in M_{1/2}(\tilde{\Gamma}_0(4))$ at the three cusps of $\Gamma_0(4)$.

§2. Eisenstein series of half integer weight for $\tilde{\Gamma}_0(4)$

Recall the functions $G_k(z)$ and $E_k(z)$ in $M_k(\Gamma)$ that we defined in §III.2 for $k \geq 4$ an even integer. We set $G_k(z) = \Sigma\,(mz+n)^{-k}$, where the sum is over $m, n \in \mathbb{Z}$ not both zero; and then $E_k(z)$ was obtained by dividing by $G_k(i\infty) = 2\zeta(k)$. Alternately, we can obtain $E_k(z)$ by the same type of summation $\Sigma\,(mz+n)^{-k}$ if we sum only over pairs m, n which have no common factor (see Problem 1 of §III.2):

$$E_k(z) = \sum \frac{1}{(mz+n)^k}, \tag{2.1}$$

where the sum is over $m, n \in \mathbb{Z}$ with g.c.d.$(m, n) = 1$ and only one pair taken from (m, n), $(-m, -n)$. In §III.2 we obtained the q-expansion

$$E_k(z) = 1 - \frac{2k}{B_k} \sum_{n=1}^{\infty} \sigma_{k-1}(n) q^n, \qquad q = e^{2\pi i z}.$$

We now want to give an analogous construction, with k replaced by a half-integer $k/2$ and Γ replaced by $\Gamma_0(4)$. To do this, we give a more intrinsic description of the sum (2.1). Notice that if we complete the row (m, n) to form a matrix $\gamma = \left(\begin{smallmatrix} a & b \\ m & n \end{smallmatrix}\right) \in \Gamma$, which can be done because g.c.d.$(m, n) = 1$, then the summand in (2.1) is the reciprocal of the automorphy factor $J_k(\gamma, z)$ for functions in $M_k(\Gamma)$. That is, such functions f satisfy: $f(\gamma z) = J_k(\gamma, z) f(z)$ for $J_k(\gamma, z) = (mz+n)^k$. In (2.1) we are summing $1/J_k(\gamma, z)$ over all equivalence classes of $\gamma \in \Gamma$, where $\gamma_1 \sim \gamma_2$ if γ_1 and γ_2 have the same bottom row up to a sign, i.e., if $\gamma_2 = \pm\left(\begin{smallmatrix} 1 & j \\ 0 & 1 \end{smallmatrix}\right)\gamma_1$ for some $\pm\left(\begin{smallmatrix} 1 & j \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma_\infty$. In other words, we may regard $J_k(\gamma, z)$ as a function on the set $\Gamma_\infty \backslash \Gamma$ of right cosets, and then rewrite (2.1) as follows:

$$E_k(z) = \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} J_k(\gamma, z)^{-1}, \qquad k \geq 4 \text{ even}. \tag{2.2}$$

It is now possible to give a proof of the invariance of $E_k(z)$ under $[\gamma_1]_k$ for $\gamma_1 \in \Gamma$ which easily generalizes to other such sums of automorphy factors. Namely, note that, by the definition of $[\gamma_1]_k$, we have

$$\begin{aligned} E_k(z)|[\gamma_1]_k &= J_k(\gamma_1, z)^{-1} E_k(\gamma_1 z) \\ &= J_k(\gamma_1, z)^{-1} \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} J_k(\gamma, \gamma_1 z)^{-1} \\ &= \sum_{\gamma \in \Gamma_\infty \backslash \Gamma} J_k(\gamma\gamma_1, z)^{-1}, \end{aligned}$$

because of the general relation (1.1) satisfied by any automorphy factor.

But right multiplication by γ_1 merely rearranges the right cosets $\Gamma_\infty\gamma$. Thus, the last sum is just a rearrangement of the sum in (2.2) for E_k. Since the sum is absolutely convergent for $k > 2$, we conclude that $E_k|[\gamma_1]_k = E_k$.

We now mimic this procedure for half integral weight $k/2$ and the group $\tilde{\Gamma}_0(4)$. In order to have absolute convergence, we must assume $k/2 > 2$, i.e., $k \geq 5$.

Note that we obviously have $\Gamma_0(4)_\infty = \Gamma_\infty = \{\pm(\begin{smallmatrix}1 & j\\ 0 & 1\end{smallmatrix})\}$.

Definition. Let k be an odd integer, $k \geq 5$.

$$E_{k/2}(z) = \sum_{\gamma\in\Gamma_\infty\backslash\Gamma_0(4)} j(\gamma, z)^{-k}. \tag{2.3}$$

As representatives γ of $\Gamma_\infty\backslash\Gamma_0(4)$ we take one matrix $(\begin{smallmatrix}a & b\\ m & n\end{smallmatrix})\in\Gamma_0(4)$ for each (m, n) with $4|m$, g.c.d.$(m, n) = 1$ (in particular, n is odd), and we may specify $n > 0$ so that we get only one of the pairs $\pm(m, n)$. Thus

$$E_{k/2}(z) = \sum_{m,n\in\mathbb{Z},4|m,n>0,\text{g.c.d.}(m,n)=1} j\left(\begin{pmatrix}a & b\\ m & n\end{pmatrix}, z\right)^{-k}.$$

Substituting the definition of $j(\gamma, z)$ and noting that $(\frac{m}{n})^{-k} = (\frac{m}{n})$, since k is odd, we obtain

$$E_{k/2}(z) = \sum_{\substack{4|m,n>0\\ \text{g.c.d.}(m,n)=1}} \left(\frac{m}{n}\right)\varepsilon_n^k(mz+n)^{-k/2}, \qquad k \geq 5 \text{ odd}. \tag{2.4}$$

The fact that $E_{k/2}\in M_{k/2}(\tilde{\Gamma}_0(4))$ now follows by the exact same argument as in the case of the Eisenstein series of integral weight for Γ which we considered above. Namely, $E_{k/2}(z)|[\tilde{\gamma}_1]_{k/2} = j(\gamma_1, z)^{-k}E_{k/2}(\gamma_1 z)$, which just gives a rearrangement of the sum (2.3). Finally, the cusp condition is routine to check, so we shall leave that as an exercise (see below).

Unlike the case of Γ, where there is only one Eisenstein series of each weight, $E_{k/2}(z)$ has a companion Eisenstein series $F_{k/2}(z)$, defined by

$$F_{k/2} = E_{k/2}\left|\left[\left(\begin{pmatrix}0 & -1\\ 4 & 0\end{pmatrix}, \sqrt{2z}\right)\right]_{k/2}\right..$$

That there are two basic Eisenstein series for each half-integer $k/2$ is related to the fact that $\Gamma_0(4)$ has two regular cusps (see Problem 3 of §1 and the discussion at the very end of §III.5). To see that $F_{k/2}\in M_{k/2}(\tilde{\Gamma}_0(4))$, we apply Problem 6 of §1 in the case $N = 4$, $\chi = 1$.

To write $F_{k/2}$ more explicitly, we compute

$$\begin{aligned} F_{k/2}(z) &= (2z)^{-k/2}E_{k/2}(-1/4z)\\ &= \frac{2^{k/2}}{(4z)^{k/2}}\sum_{\substack{4|m,n>0\\ \text{g.c.d.}(m,n)=1}}\left(\frac{m}{n}\right)\varepsilon_n^k\frac{1}{(-m/4z+n)^{k/2}}\\ &= 2^{k/2}\sum\left(\frac{m}{n}\right)\varepsilon_n^k\frac{1}{(-m+4nz)^{k/2}}, \end{aligned}$$

where we have used the fact that $n > 0$ to write $\sqrt{4z}\sqrt{-m/4z + n} = \sqrt{-m + 4nz}$. We now replace m by $-4m$, so that the sum becomes a summation over all $m \in \mathbb{Z}$ with g.c.d.$(4m, n) = 1$, i.e., n odd and g.c.d.$(m, n) = 1$. We obtain

$$F_{k/2}(z) = 2^{-k/2} \sum_{\substack{m,n \in \mathbb{Z} \\ n>0 \text{ odd} \\ \text{g.c.d.}(m,n)=1}} \left(\frac{-m}{n}\right) \varepsilon_n^k \frac{1}{(nz+m)^{k/2}}. \tag{2.5}$$

We next compute the q-expansions of $E_{k/2}$ and $F_{k/2}$. We proceed much as in the computation of the q-expansion of E_k in §III.2, except that there are added complications.

The derivation of the q-expansion for $F_{k/2}(z)$ turns out to be slightly less tedious than for $E_{k/2}(z)$. Hence, we shall give that derivation, and omit the analogous proof of the q-expansion formula for $E_{k/2}(z)$.

To compute the q-expansion of $F_{k/2}(z)$, we start with the following relation, which was proved in Problem 8(c) of §II.4 for any $a > 1$, $z \in H$ (where we have replaced a, s by z, a here):

$$\sum_{h=-\infty}^{\infty} (z+h)^{-a} = (2\pi)^a e^{-\pi i a/2} \Gamma(a)^{-1} \sum_{l=1}^{\infty} l^{a-1} e^{2\pi i l z}. \tag{2.6}$$

(Note: for $z \in H$ we define $z^a = e^{a \log z}$, where $\log z$ denotes the branch having imaginary part in $(0, \pi)$ for $z \in H$.)

Since the sum (2.5) for $F_{k/2}(z)$ is absolutely convergent, we may order the terms as we please. If we let $j = 0, 1, \ldots, n-1$ and write $m = -j + nh$, $h \in \mathbb{Z}$, we obtain

$$\begin{aligned} F_{k/2}(z) &= 2^{-k/2} \sum_{n>0 \text{ odd}} \varepsilon_n^k \sum_{0 \le j < n} \left(\frac{j}{n}\right) \sum_{h=-\infty}^{\infty} (nz - j + nh)^{-k/2} \\ &= 2^{-k/2} \sum_{n>0 \text{ odd}} \varepsilon_n^k n^{-k/2} \sum_{0 \le j < n} \left(\frac{j}{n}\right) (2\pi)^{k/2} e^{-\pi i k/4} \Gamma\left(\frac{k}{2}\right)^{-1} \sum_{l=1}^{\infty} l^{(k/2)-1} e^{-2\pi i l j/n} e^{2\pi i l z}. \end{aligned}$$

by (2.6) with z replaced by $z - j/n$ and a replaced by $k/2$.

Thus, $F_{k/2}(z) = \sum_{l=1}^{\infty} b_l q^l$, $q = e^{2\pi i z}$, where

$$b_l = \frac{\pi^{k/2}}{\Gamma(\frac{k}{2}) e^{\pi i k/4}} l^{(k/2)-1} \sum_{n>0 \text{ odd}} \varepsilon_n^k n^{-k/2} \sum_{0 \le j < n} \left(\frac{j}{n}\right) e^{-2\pi i l j/n}. \tag{2.7}$$

We now simplify the expression (2.7) in the important special case when l is squarefree.

Let $m \in \mathbb{Z}$ be squarefree. For $j > 0$ odd we let $\chi_m(j)$ denote $(\frac{m}{j})$. Thus, if j is a prime, it is $\chi_m(j)$ which tells us whether j splits, remains prime, or ramifies in $\mathbb{Q}(\sqrt{m})$ (i.e., this depends on whether $\chi_m(j)$ equals 1, -1, or 0, respectively). There is a unique character, also denoted χ_m, which has conductor $|m|$ if $m \equiv 1 \bmod 4$ and conductor $|4m|$ if $m \equiv 2$ or $3 \bmod 4$ and which agrees with $\chi_m(j) = (\frac{m}{j})$ for $j > 0$ odd. We can express χ_m in terms of the usual Legendre symbol as follows:

$$\chi_m(j) = \left(\frac{j}{|m|}\right) \qquad \text{if } m \equiv 1 \bmod 4;$$

$$\chi_m(j) = \left(\frac{-1}{j}\right)\left(\frac{j}{|m|}\right) \qquad \text{if } m \equiv -1 \bmod 4;$$

$$\chi_m(j) = \left(\frac{2}{j}\right)\chi_{m_0}(j) \qquad \text{if } m = 2m_0 \equiv 2 \bmod 4,$$

where we define $(\frac{-1}{j}) = (-1)^{(j-1)/2}$ for j odd and $(\frac{-1}{j}) = 0$ for j even, and $(\frac{2}{j}) = (-1)^{(j^2-1)/8}$ for j odd and $(\frac{2}{j}) = 0$ for j even.

Note that

$$\chi_m(-1) = 1 \quad \text{if} \quad m > 0, \quad \text{and} \quad \chi_m(-1) = -1 \quad \text{if} \quad m < 0. \tag{2.8}$$

Recall that we define $\lambda = (k-1)/2$.

Proposition 5. *The Eisenstein series $F_{k/2}$ has q-expansion coefficient b_l which for l squarefree is equal to $L(\chi_{(-1)^\lambda l}, 1-\lambda)$ times a factor that depends only on λ and on l* mod 8, *but not on l itself. This factor is given in* (2.16) *below.*

We first prove a lemma.

Lemma. *Let $n = n_0 n_1^2$, where n_0 is squarefree. Then for l squarefree, the inner sum in* (2.7) *vanishes unless $n_1 | l$, and if $n_1 | l$ it equals*

$$\varepsilon_n \left(\frac{-l}{n_0}\right)\sqrt{n_0}\,\mu(n_1)n_1, \tag{2.9}$$

where μ is the Möbius function (equal to 0 *if n_1 is not squarefree and equal to $(-1)^r$ if n_1 is the product of r distinct primes).*

Proof of Lemma. First, if $n_1 = 1$, then $(\frac{j}{n})$ has conductor $n = n_0$. If l has a common factor d with n, then, combining terms which differ by multiples of n/d, we see that the inner sum in (2.7) is zero. On the other hand, if $l \in (\mathbb{Z}/n\mathbb{Z})^*$, then making the change of variables $j' = -lj$ leads to $(\frac{-l}{n})\Sigma(\frac{j'}{n})e^{2\pi i j'/n}$. This Gauss sum is well known to equal $\varepsilon_n\sqrt{n}$ (see, e.g., §5.2 and §5.4 of [Borevich and Shafarevich 1966]). This proves the lemma when $n_1 = 1$.

We next show that the inner sum is zero if g.c.d.$(n_0, n_1) = d > 1$. Note that

$$\left(\frac{j}{n}\right) = \begin{cases} \left(\frac{j}{n_0}\right) & \text{if g.c.d.}(j, n) = 1, \\ 0 & \text{otherwise;} \end{cases}$$

and it follows that $(\frac{j}{n}) = (\frac{j}{n/d^2})$. Combining terms with j which differ by multiples of n/d^2 and using the fact that $d^2 \nmid l$ (since l is squarefree), so that $\Sigma_{j'} e^{-2\pi i l(j+j'n/d^2)/n} = 0$, we find that the inner sum is zero in this case.

Now suppose that $n_1 > 1$ is prime to n_0. Let p_ν run through all primes dividing n_1: $n_1 = \prod p_\nu^{a_\nu}$. Set $q_\nu = p_\nu^{a_\nu}$; and let $\delta_\nu(j) = 0$ if $p_\nu | j$ and $\delta_\nu(j) = 1$ if $p_\nu \nmid j$. Then $(\frac{j}{n}) = (\frac{j}{n_0}) \prod_\nu \delta_\nu(j)$. Any $j \in \mathbb{Z}/n\mathbb{Z}$ can be written uniquely in the form

$$j = j_0 n_1^2 + \sum j_\nu n/q_\nu^2, \qquad 0 \le j_0 < n_0, \qquad 0 \le j_\nu < q_\nu^2.$$

Then

$$e^{-2\pi i l j/n} = e^{-2\pi i l j_0/n_0} \prod_\nu e^{-2\pi i l j_\nu/q_\nu^2}.$$

Further note that

$$\left(\frac{j}{n}\right) = \left(\frac{j}{n_0}\right) \prod_\nu \delta_\nu(j) = \left(\frac{j_0}{n_0}\right) \prod_\nu \delta_\nu(j_\nu).$$

Thus, the inner sum in (2.7) becomes

$$\left(\sum_{j_0} \left(\frac{j_0}{n_0}\right) e^{-2\pi i l j_0/n_0}\right) \prod_\nu \sum_{j_\nu} \delta_\nu(j_\nu) e^{-2\pi i l j_\nu/q_\nu^2}. \tag{2.10}$$

The first sum is $\varepsilon_{n_0}(\frac{-l}{n_0})\sqrt{n_0}$, by the first part of the proof (the case $n = n_0$). The sum inside the product is

$$\sum_{j_\nu=0}^{q_\nu^2-1} e^{-2\pi i l j_\nu/q_\nu^2} - \sum_{j_\nu=0}^{q_\nu^2/p_\nu-1} e^{-2\pi i l p_\nu j_\nu/q_\nu^2}.$$

Here the first sum is zero, because $q_\nu^2 \nmid l$. The second sum is also zero, unless $q_\nu^2 | l p_\nu$; this can happen only if $q_\nu = p_\nu$ and $p_\nu | l$. Thus, the expression (2.10) is zero unless each $q_\nu = p_\nu$, i.e., n_1 is squarefree, and $n_1 | l$. If $n_1 | l$, then each sum inside the product in (2.10) is equal to $-p_\nu$, and we obtain the following expression for the desired sum:

$$\varepsilon_n \left(\frac{-l}{n_0}\right) \sqrt{n_0} \prod_\nu (-p_\nu).$$

(Note that $\varepsilon_{n_0} = \varepsilon_n$.) The latter product is $\mu(n_1)n_1$. This completes the proof of the lemma. □

We are now ready to compute b_l for l squarefree.

We replace the inner sum in (2.7) by its value given in the lemma, and sum over all n with a fixed n_0. According to the lemma, we obtain

$$b_l = \frac{\pi^{k/2} l^{k/2-1}}{\Gamma(\frac{k}{2}) e^{\pi i k/4}} \sum_{\substack{n_0 > 0 \text{ odd} \\ \text{and squarefree}}} \sum_{\substack{n_1 | l \\ n_1 \text{ odd}}} \varepsilon_{n_0 n_1^2}^{k+1} (n_0 n_1^2)^{-k/2} \left(\frac{-l}{n_0}\right) \sqrt{n_0} \mu(n_1) n_1.$$

Since $\varepsilon_{n_0 n_1^2} = \varepsilon_{n_0}$, the expression inside the summation over n_0 is equal to

$$\varepsilon_{n_0}^{k+1} \left(\frac{-l}{n_0}\right) n_0^{(1-k)/2} \sum_{\text{odd } n_1 | l} \mu(n_1) n_1^{1-k}.$$

The sum over $n_1 | l$ is equal to $\prod_{\text{odd } p | l}(1 - p^{1-k})$; since it does not depend on

n_0, we can pull it outside the summation over n_0. In addition, we use the relation (recall: $\lambda = (k-1)/2$)

$$\varepsilon_{n_0}^{k+1}\left(\frac{-l}{n_0}\right) = \left(\frac{-1}{n_0}\right)^{(k+1)/2}\left(\frac{-1}{n_0}\right)\left(\frac{l}{n_0}\right) = \left(\frac{-1}{n_0}\right)^{\lambda}\left(\frac{l}{n_0}\right).$$

As a result, we obtain

$$b_l = \frac{\pi^{k/2} l^{k/2-1}}{\Gamma(\frac{k}{2})e^{\pi i k/4}} \prod_{\substack{p|l\\ p\,\mathrm{odd}}} (1-p^{-2\lambda}) \sum_{\substack{n_0>0\,\mathrm{odd}\\ \mathrm{and\ squarefree}}} \left(\frac{-1}{n_0}\right)^{\lambda}\left(\frac{l}{n_0}\right) n_0^{-\lambda}. \qquad (2.11)$$

The expression $(\frac{-1}{n_0})^{\lambda}(\frac{l}{n_0})$ is what we denoted $\chi_{(-1)^\lambda l}(n_0)$. As remarked above, these are the values at odd squarefree positive n_0 of a primitive character $\chi_{(-1)^\lambda l}$ of conductor $N = l$ if $(-1)^\lambda l \equiv 1 \bmod 4$ and conductor $N = 4l$ if $(-1)^\lambda l \equiv 2$ or $3 \bmod 4$.

Thus, the sum in (2.11) looks very much like the value at $s = \lambda$ of the Dirichlet L-series

$$L(\chi_{(-1)^\lambda l}, s) \underset{\mathrm{def}}{=} \sum_{n=1}^{\infty} \chi_{(-1)^\lambda l}(n) n^{-s}, \qquad \mathrm{Re}\, s > 1. \qquad (2.12)$$

The difference between the two sums is that the sum in (2.11) only ranges over all odd squarefree n. But if we write an arbitrary n in the form $n = n_0 n_1^2$ with n_0 squarefree, we see that $\chi_{(-1)^\lambda l}(n) = \chi_{(-1)^\lambda l}(n_0)$ if n_1 is prime to N, and $\chi_{(-1)^\lambda l}(n) = 0$ if g.c.d.$(n, N) > 1$, where $N = l$ or $4l$ is the conductor.

First suppose that $N = 4l$. Then, if we were to multiply the sum in (2.11) by

$$\sum_{n_1>0\ \mathrm{prime\ to}\ N} (n_1^2)^{-\lambda} = \prod_{p\nmid N} (1-p^{-2\lambda})^{-1} = \zeta(2\lambda)\prod_{p|N}(1-p^{-2\lambda}),$$

we would obtain the L-series (2.12) with $s = \lambda$. In other words, we have

$$b_l \zeta(2\lambda) \prod_{p|N} (1-p^{-2\lambda}) = \frac{\pi^{k/2} l^{k/2-1}}{\Gamma(\frac{k}{2})e^{\pi i k/4}} L(\chi_{(-1)^\lambda l}, \lambda) \prod_{\substack{p|l\\ p\neq 2}} (1-p^{-2\lambda}).$$

The products cancel, except for $p = 2|N$ on the left, so we obtain (here we replace k by $2\lambda + 1$):

$$b_l = \frac{L(\chi_{(-1)^\lambda l}, \lambda)}{\zeta(2\lambda)} \frac{l^{\lambda-1/2}}{1-2^{-2\lambda}} \frac{\pi^{\lambda+1/2}}{\Gamma(\lambda+\frac{1}{2})e^{\pi i(\lambda/2+1/4)}} \qquad (2.13)$$
$$\text{for} \quad (-1)^\lambda l \equiv 2 \text{ or } 3 \bmod 4.$$

If, on the other hand, $N = l$, i.e., $(-1)^\lambda l \equiv 1 \bmod 4$, then to obtain $L(\chi_{(-1)^\lambda l}, \lambda)$ we must multiply the sum in (2.11) by

$$\frac{1}{1-\chi_{(-1)^\lambda l}(2)2^{-\lambda}} \prod_{p\nmid 2l} \frac{1}{1-p^{-2\lambda}},$$

where the first term $(1-\chi_{(-1)^\lambda l}(2)2^{-\lambda})^{-1}$ is necessary in order to get the terms in (2.12) with n even. Since $1-2^{-2\lambda} = (1+\chi_{(-1)^\lambda l}(2)2^{-\lambda})(1-\chi_{(-1)^\lambda l}(2)2^{-\lambda})$,

we can rewrite the last product as

$$(1+\chi_{(-1)^\lambda l}(2)2^{-\lambda})\prod_{p\nmid l}(1-p^{-2\lambda})^{-1}.$$

This leads to the equality

$$b_l = \frac{L(\chi_{(-1)^\lambda l}, \lambda)}{\zeta(2\lambda)} \frac{l^{\lambda-1/2}}{1+\chi_{(-1)^\lambda l}(2)2^{-\lambda}} \frac{\pi^{\lambda+1/2}}{\Gamma(\lambda+\frac{1}{2})e^{\pi i(\lambda/2+1/4)}} \quad \text{for} \quad (-1)^\lambda l \equiv 1 \bmod 4. \tag{2.14}$$

The formulas (2.13) and (2.14) express the l-th q-expansion coefficient of $F_{k/2}$ in terms of the values at positive integers of an L-function and the Riemann zeta-function. Although these formulas include complicated-looking factors, these factors largely disappear if we use the functional equations for $L(\chi_{(-1)^\lambda l}, s)$ and $\zeta(s)$ to express b_l in terms of values at the *negative* integers.

Namely, by the theorem in §II.4 we have

$$\zeta(2\lambda) = \zeta(1-2\lambda)\pi^{2\lambda-1/2}\Gamma(\tfrac{1}{2}-\lambda)/\Gamma(\lambda).$$

Next, Problems 3(e) and 5(e) of §II.4 give us the functional equation for $L(\chi_{(-1)^\lambda l}, s)$, where the functional equation is slightly different for even and odd characters. By (2.8) it follows that $\chi_{(-1)^\lambda l}$ is an even character if λ is even and an odd character if λ is odd. We thus have

$$L(\chi_{(-1)^\lambda l}, \lambda) = \left(\frac{\pi}{N}\right)^{\lambda-(1/2)} L(\chi_{(-1)^\lambda l}, 1-\lambda) \cdot \begin{cases} \Gamma(\frac{1}{2}-\frac{1}{2}\lambda)/\Gamma(\frac{1}{2}\lambda) & \text{if } \lambda \text{ is even;} \\ \Gamma(1-\frac{1}{2}\lambda)/\Gamma(\frac{1}{2}+\frac{1}{2}\lambda) & \text{if } \lambda \text{ is odd.} \end{cases} \tag{2.15}$$

Substituting these equalities in (2.14) and using the relations $\Gamma(\lambda) = \Gamma(\frac{1}{2}\lambda)\Gamma(\frac{1}{2}+\frac{1}{2}\lambda)2^{\lambda-1}/\sqrt{\pi}$ (see (4.4) of §II.4) and $\Gamma(x)\Gamma(1-x) = \pi/\sin \pi x$ (see (4.3) of §II.4), one obtains (the details will be left to the reader):

$$b_l = \frac{L(\chi_{(-1)^\lambda l}, 1-\lambda)}{(1+(-1)^\lambda i)\zeta(1-2\lambda)} \cdot \begin{cases} \dfrac{2^{\lambda-1/2}}{1+\chi_{(-1)^\lambda l}(2)2^{-\lambda}} & \text{if } (-1)^\lambda l \equiv 1 \bmod 4; \\ \dfrac{2^{1/2-\lambda}}{1-2^{-2\lambda}} & \text{if } (-1)^\lambda l \equiv 2 \text{ or } 3 \bmod 4. \end{cases} \tag{2.16}$$

Notice that the L-function value depends on the precise value of l, but all of the other factors are either independent of l or else only depend on the value of l modulo 8. This completes the proof of Proposition 5. □

For fixed λ even,as l ranges through squarefree positive integers, the even characters $\chi_{(-1)^\lambda l}$ run through the characters of all real quadratic fields $\mathbb{Q}(\sqrt{l})$. The discriminant D of $\mathbb{Q}(\sqrt{l})$ is $D = l$ if $l \equiv 1 \bmod 4$ and $D = 4l$

if $l \equiv 2$ or 3 mod 4. For fixed λ odd, the odd characters $\chi_{(-1)^\lambda l}$ run through the characters corresponding to all imaginary quadratic fields $\mathbb{Q}(\sqrt{-l})$, whose discriminant is $D = -l$ if $-l \equiv 1$ mod 4 and $D = -4l$ if $-l \equiv 2$ or 3 mod 4.

Thus, for $(-1)^\lambda l \equiv 1$ mod 4, the formula (2.16) expresses the $|D|$-th q-expansion coefficient of $F_{\lambda+1/2}$ in terms of the value at $1-\lambda$ of the L-function of the quadratic character of conductor $|D|$.

By a similar computation, it is not hard to show that for $(-1)^\lambda l \equiv 2$ or 3 mod 4, $|D| = 4l$, the $|D|$-th q-expansion coefficient is given by

$$b_{|D|} = b_{4l} = 2^{2\lambda-1}b_l = \frac{L(\chi_{(-1)^\lambda l}, 1-\lambda)}{(1+(-1)^\lambda i)\zeta(1-2\lambda)} \cdot \frac{2^{\lambda-1/2}}{1-2^{-2\lambda}}. \tag{2.17}$$

We now look at the q-expansion coefficients of the other Eisenstein series $E_{\lambda+1/2}$. By an argument similar to the derivation of (2.16)–(2.17), one shows that for l squarefree the l-th coefficient a_l in the q-expansion of $E_{k/2} = \Sigma\, a_l q^l$ is given by

$$a_l = (1+(-1)^\lambda i)2^{-1/2-\lambda}b_l \cdot \begin{cases} 1 + \chi_{(-1)^\lambda l}(2)2^{1-\lambda} & \text{if } (-1)^\lambda l \equiv 1 \bmod 4; \\ -1 & \text{if } (-1)^\lambda l \equiv 2 \text{ or } 3 \bmod 4 \end{cases} \tag{2.18}$$

$$= \frac{L(\chi_{(-1)^\lambda l}, 1-\lambda)}{\zeta(1-2\lambda)} \cdot \begin{cases} \dfrac{1}{2} \cdot \dfrac{1 + \chi_{(-1)^\lambda l}(2)2^{1-\lambda}}{1 + \chi_{(-1)^\lambda l}(2)2^{-\lambda}} & \text{if } (-1)^\lambda l \equiv 1 \bmod 4; \\ \dfrac{1}{1-2^{2\lambda}} & \text{if } (-1)^\lambda l \equiv 2 \text{ or } 3 \bmod 4; \end{cases} \tag{2.19}$$

and that when $(-1)^\lambda l \equiv 2$ or 3 mod 4, so that $D = (-1)^\lambda 4l$ is the discriminant of a quadratic field, we have

$$a_{4l} = (1+(-1)^\lambda i)2^{-(1/2)-\lambda}(1-2^{1-2\lambda})b_{4l} = \frac{L(\chi_{(-1)^\lambda l}, 1-\lambda)}{\zeta(1-2\lambda)} \cdot \frac{1-2^{1-2\lambda}}{2-2^{1-2\lambda}}. \tag{2.20}$$

Thus, we have found that both $E_{k/2}$ and $F_{k/2}$ have $|D|$-th q-expansion coefficient equal to $L(\chi_D, 1-\lambda)$ times a factor which depends only on λ and, in the case $D = (-1)^\lambda l \equiv 1$ mod 4, on $\chi_D(2)$, which equals 1 if $(-1)^\lambda l \equiv 1$ mod 8 and -1 if $(-1)^\lambda l \equiv 5$ mod 8. For now, let us abbreviate $\chi = \chi_D(2)$.

The same is true of any linear combination of $E_{k/2}$ and $F_{k/2}$, so it should be possible to find a linear combination whose $|D|$-th q-expansion coefficient is *exactly* $L(\chi_D, 1-\lambda)$. More precisely, we look for coefficients α and β such that

$$\zeta(1-2\lambda)(\alpha E_{k/2} + \beta F_{k/2})$$

has $|D|$-th q-expansion coefficient $c_{|D|}$ equal to $L(\chi_D, 1-\lambda)$ for all discriminants D of real quadratic fields if λ is even and of imaginary quadratic

fields if λ is odd. The term $\zeta(1-2\lambda)$ is inserted for simplicity, in order to cancel the term $\zeta(1-2\lambda)$ in the denominators of all q-expansion coefficients of $E_{k/2}$ and $F_{k/2}$. Thus, we want:

for $D = (-1)^\lambda l \equiv 1 \bmod 4$,

$$L(\chi_D, 1-\lambda) = c_{|D|}$$
$$= L(\chi_D, 1-\lambda)\left(\frac{1}{2}\cdot\frac{1+2^{1-\lambda}\chi}{1+2^{-\lambda}\chi}\alpha + \frac{1}{1+(-1)^\lambda i}\cdot\frac{2^{\lambda-(1/2)}}{1+2^{-\lambda}\chi}\beta\right);$$

for $D = (-1)^\lambda 4l$, $(-1)^\lambda l \equiv 2$ or $3 \bmod 4$,

$$L(\chi_D, 1-\lambda) = c_{|D|}$$
$$= L(\chi_D, 1-\lambda)\left(\frac{1}{2}\cdot\frac{1-2^{1-2\lambda}}{1-2^{-2\lambda}}\alpha + \frac{1}{1+(-1)^\lambda i}\cdot\frac{2^{\lambda-(1/2)}}{1-2^{-2\lambda}}\beta\right).$$

Dividing by $L(\chi_D, 1-\lambda)$, we obtain a 2×2 system of equations in the unknowns α, β. Because χ depends on l modulo 8, it is not *a priori* clear that the α and β for which $\alpha a_{|D|} + \beta b_{|D|} = L(\chi_D, 1-\lambda)/\zeta(1-2\lambda)$ will be independent of l. However, when we solve the equations for α and β, we find that the solution does not actually contain χ, and so is independent of l. Namely, we obtain

$$\alpha = 1, \qquad \beta = (1+(-1)^\lambda i)2^{-\lambda-(1/2)}.$$

We conclude

Proposition 6. *Let $\lambda = (k-1)/2 \geq 2$, and let $E_{k/2}, F_{k/2} \in M_{k/2}(\tilde{\Gamma}_0(4))$ be defined by* (2.4)–(2.5). *Then*

$$H_{k/2} \underset{\text{def}}{=} \zeta(1-2\lambda)(E_{k/2} + (1+i^k)2^{-k/2}F_{k/2}) \in M_{k/2}(\tilde{\Gamma}_0(4))$$

has the following property: if $D = (-1)^\lambda l$ or $(-1)^\lambda 4l$, $l > 0$, is the discriminant of a quadratic field and χ_D is the corresponding character, then the $|D|$-th q-expansion coefficient of $H_{k/2}$ is equal to $L(\chi_D, 1-\lambda)$.

This proposition is due to H. Cohen [1975]. (His notation is slightly different from ours.) It can be viewed as a prototype for the theorem of Waldspurger–Tunnell which we shall discuss later. The Waldspurger–Tunnell theorem exhibits a modular form of weight $\frac{3}{2}$ (think of λ as being equal to 1 in Proposition 6) for $\Gamma_0(128)$ such that the square of its n-th q-expansion coefficient is a certain nonzero factor times $L(E_n, 1)$, where $L(E, s)$ is the Hasse–Weil L-function of an elliptic curve E (see §II.5) and E_n is the elliptic curve $y^2 = x^3 - n^2x$. Thus, the q-expansion coefficients of this particular form of half integral weight are closely related to all of the critical values $L(E_n, 1)$ which we encountered in Chapter II.

If we were allowed to take $\lambda = 1$ in Proposition 6, then the $|D|$-th q-expansion coefficient would be $L(\chi_D, 0)$, which differs by a simple nonzero

factor from $L(\chi_D, 1)$ (because of the functional equation), and is essentially equal to the class number of the quadratic imaginary field $\mathbb{Q}(\sqrt{D})$ (here $D = -l$ or $-4l$). We will say more about the case $\lambda = 1$ below.

For another proof of Proposition 6, using so-called "Jacobi forms," see [Eichler and Zagier 1984].

We now discuss some interesting consequences of Proposition 6. Since any element $H_{k/2} \in M_{k/2}(\tilde{\Gamma}_0(4))$ can be expressed as a polynomial in $\Theta = \Sigma_{n=-\infty}^{\infty} q^{n^2}$ and $F = \Sigma_{n>0 \text{ odd}} \sigma_1(n) q^n$ (see Proposition 4), it follows that there are formulas expressing $L(\chi_D, 1-\lambda)$ in terms of $\sigma_1(n)$ and the number of ways n can be expressed as a sum of r squares. We illustrate with the simple case $k = 5$, $\lambda = 2$. In this case $M_{5/2}(\tilde{\Gamma}_0(4))$ is spanned by Θ^5 and ΘF, and a comparison of the constant coefficients and the coefficients of the first power of q shows that $H_{5/2} = \frac{1}{120}\Theta^5 - \frac{1}{6}\Theta F$ (see the exercises below). This yields the following proposition.

Proposition 7. *Let $s_r(n)$ denote the number of ways n can be written as a sum of r squares; thus $\Theta^5 = \Sigma s_5(n) q^n$. Let D be the discriminant of a real quadratic field, and let χ_D be the corresponding character. Then*

$$L(\chi_D, -1) = \frac{1}{120} s_5(D) - \frac{1}{6} \sum_{\substack{|j| < \sqrt{D} \\ D - j^2 \text{ odd}}} \sigma_1(D - j^2).$$

Many other relations of this sort can be derived, in much the same way as we used the structure of $M_k(\Gamma)$ in the exercises in §III.2 to derive identities between various $\sigma_r(n)$. For details, see [Cohen 1975]. These relations can be viewed as a generalization to $\lambda > 1$ of the so-called "class number relations" of Kronecker and Hurwitz. (For a statement and proof of the class number relations, see, for example, Zagier's appendix to [Lang 1976] and his correction following the article [Zagier 1977].) The class number relations correspond to the case $\lambda = 1$, i.e., $L(\chi_D, 1-\lambda) = L(\chi_D, 0)$. But $\lambda = 1$ falls outside the range of applicability of Proposition 6.

To get at the class number relations from the point of view of forms of half integral weight, one must study a more complicated function $H_{3/2}$, which transforms under $\Gamma_0(4)$ like a form of weight $\frac{3}{2}$ but which is not analytic. The $|D|$-th coefficient of $H_{3/2}$ is essentially the class number of the imaginary quadratic field $\mathbb{Q}(\sqrt{D})$. The study of $H_{3/2}$ is due to Zagier [1975a].

In some sense, the situation with $H_{3/2}$ is analogous to the situation with E_2 in §III.2. In both cases, when we lower the weight just below the minimum weight necessary for absolute convergence, problems arise and we no longer have a true modular form. The study of E_2 and $H_{3/2}$ is much subtler than that of E_k and $H_{k/2}$ for $k > 2$ and $k/2 > 2$, respectively. But in both cases the Eisenstein series, though they fail to be modular forms, have a special importance: we saw that the functional equation for E_2 led to the functional equation for the discriminant form $\Delta(z) \in S_{12}(\Gamma)$; and $H_{3/2}$ has as its coefficients the class numbers of all imaginary quadratic fields.

We next examine the quite different question of how the *non*squarefree q-expansion coefficients of $E_{k/2}$, $F_{k/2}$, and $H_{k/2}$ can be determined. Again we shall give the details only for $F_{k/2}$ and shall omit the similar computations for $E_{k/2}$.

We proceed one prime at a time. That is, for each prime p we express b_l in terms of b_{l_0}, where $l = p^{2v}l_0$ and $p^2 \nmid l_0$. Finally, combining the different primes p, we shall be able to express b_l in terms of b_{l_0} when $l = l_0 l_1^2$ and l_0 is squarefree. This information can be conveniently summarized in the form of an Euler product for $\Sigma_{l_1=1}^{\infty} b_{l_0 l_1^2} l_1^{-s}$, as we shall see later.

The easiest case is $p = 2$. Suppose that $l = 4^v l_0$, $4 \nmid l_0$. We return to our earlier computation of b_l in (2.7). The general formula (2.7) is valid for all l, not necessarily squarefree. If we replace l by $4l$ in (2.7), the double sum does not change, because replacing l by $4l$ is equivalent to replacing j in the inner sum by $4j$, which runs through $\mathbb{Z}/n\mathbb{Z}$ as j does, since n is odd. Thus, $b_{4l} = 4^{k/2-1} b_l$, and so for $l = 4^v l_0$ we have

$$b_l = 2^{(k-2)v} b_{l_0}. \tag{2.21}$$

This case $p = 2$ can be summarized as follows: for any l_0,

$$\sum_{v=0}^{\infty} b_{l_0 4^v} 2^{-sv} = b_{l_0} \frac{1}{1 - 2^{k-2-s}}. \tag{2.22}$$

Comparing coefficients of 2^{-sv} on both sides of (2.22) shows that (2.22) is equivalent to (2.21). (Note that we do not actually need to assume that $4 \nmid l_0$.)

Now suppose that p is odd. Let $l = p^{2v} l_0$, where $p^2 \nmid l_0$. We again use (2.7) and divide into two cases.

Case (i) $p | l_0$. Let $l_0 = p\tilde{l}_0$.

We write the double sum in (2.7) in the form $\Sigma_{n>0 \text{ odd}} S_n$, where S_n is the term in (2.7) corresponding to n. Let $n = n_0 p^{2h}$, where $p^2 \nmid n_0$. We fix n_0 and look at the sum of the S_n over all $n = n_0 p^{2h}$, $h = 0, 1, 2, \ldots$. Note that $\varepsilon_n = \varepsilon_{n_0}$ and $n^{-k/2} = n_0^{-k/2} p^{-hk}$. We now evaluate the inner sum over $j \in \mathbb{Z}/n\mathbb{Z}$.

First suppose that $p \nmid n_0$. As representatives j of $\mathbb{Z}$ mod $n\mathbb{Z}$ we choose $j = j_0 p^{2h} + j_1 n_0$ as j_0 ranges from 0 to $n_0 - 1$ and j_1 ranges from 0 to $p^{2h} - 1$. With p fixed we introduce the notation

$$\delta(j) = \begin{cases} 0 & \text{if } p | j; \\ 1 & \text{if } p \nmid j. \end{cases} \tag{2.23}$$

Then for $h \geq 1$ we have

$$\left(\frac{j}{n}\right) = \left(\frac{j}{p^{2h}}\right)\left(\frac{j}{n_0}\right) = \delta(j)\left(\frac{j}{n_0}\right) = \delta(j_1)\left(\frac{j_0}{n_0}\right).$$

Also,

$$e^{-2\pi i j l/n} = e^{-2\pi i j_0 l/n_0} e^{-2\pi i j_1 \tilde{l}_0 p^{1+2(v-h)}}.$$

Thus, the term S_n in (2.7) corresponding to $n = n_0 p^{2h}$ is the product of

$$\varepsilon_{n_0}^k n_0^{-k/2} \sum_{0 \le j_0 < n_0} \left(\frac{j_0}{n_0}\right) e^{-2\pi i j_0 l/n_0} = S_{n_0}$$

and

$$p^{-hk} \sum_{0 \le j_1 < p^{2h}} \delta(j_1) e^{-2\pi i j_1 \tilde{l}_0 p^{1+2(v-h)}}.$$

In the last sum, make a change of variables by replacing $-j_1 \tilde{l}_0$ by j_1; then the sum becomes

$$\sum_{0 \le j_1 < p^{2h}} e^{2\pi i j_1 p^{1+2(v-h)}} - \sum_{0 \le j_2 < p^{2h-1}} e^{2\pi i j_2 p^{2+2(v-h)}}$$

(where $j_1 = pj_2$). The first sum is zero if $h > v$ and it is p^{2h} if $h \le v$; the second sum is zero if $h > v + 1$ and it is p^{2h-1} if $h \le v + 1$.

We conclude that all of the terms in the outer sum in (2.7) corresponding to $n = n_0 p^{2h}$ for fixed n_0 contribute

$$S_{n_0}\left(1 + \sum_{h=1}^{v} p^{-hk}(p^{2h} - p^{2h-1}) - p^{-(v+1)k} p^{2(v+1)-1}\right)$$

$$= S_{n_0}(1 - p^{1-k}) \sum_{h=0}^{v} p^{h(2-k)}.$$

We next suppose that $p|n_0$, $n_0 = p\tilde{n}_0$. As representatives j of $\mathbb{Z}$ mod $n\mathbb{Z}$ we choose

$$j = j_0 p^{2h+1} + j_1 \tilde{n}_0, \qquad 0 \le j_0 < \tilde{n}_0, \qquad 0 \le j_1 < p^{2h+1}.$$

We have

$$\left(\frac{j}{n}\right) = \left(\frac{j}{p}\right)\left(\frac{j}{\tilde{n}_0}\right) = \left(\frac{j_1}{p}\right)\left(\frac{j_0 p}{\tilde{n}_0}\right).$$

Then S_n is the product of

$$\varepsilon_n^k n^{-k/2} \sum_{0 \le j_0 < \tilde{n}_0} \left(\frac{j_0 p}{\tilde{n}_0}\right) e^{-2\pi i j_0 l/\tilde{n}_0}$$

and

$$\sum_{0 \le j_1 < p^{2h+1}} \left(\frac{j_1}{p}\right) e^{-2\pi i j_1 \tilde{l}_0 p^{2(v-h)}}.$$

But it is easy to see that for *any* h the latter sum is zero. Namely, if $h \le v$, then we obtain $\Sigma_{j_1} (\frac{j_1}{p}) = 0$; and if $h > v$, then the sum over any given residue class mod p, i.e., $j_1 = j_2 + pj_3$ for fixed j_2, is equal to

$$\left(\frac{j_2}{p}\right) e^{-2\pi i j_2 \tilde{l}_0 p^{2(v-h)}} \sum_{0 \le j_3 < p^{2h}} e^{-2\pi i j_3 \tilde{l}_0 p^{2(v-h)+1}},$$

and the latter sum is zero, because $2(v - h) + 1 < 0$ and $p \nmid \tilde{l}_0$. Thus, $S_n = 0$ if n is divisible by an odd power of p.

We conclude that for $p|l_0$ we have

$$b_l = \frac{\pi^{k/2}}{\Gamma\left(\frac{k}{2}\right)e^{\pi ik/4}} l^{(k/2)-1} \sum_{n_0>0 \text{ odd}, p\nmid n_0} S_{n_0}(1-p^{1-k}) \sum_{h=0}^{v} p^{h(2-k)}.$$

Here S_{n_0} depends only on l_0 and not on $l = l_0 p^{2v}$, because for $p\nmid n_0$

$$\sum_{0\le j<n_0} \left(\frac{j}{n_0}\right) e^{-2\pi ilj/n_0} = \sum_{0\le j<n_0} \left(\frac{jp^{2v}}{n_0}\right) e^{-2\pi il_0 jp^{2v}/n_0} = \sum_{0\le j<n_0} \left(\frac{j}{n_0}\right) e^{-2\pi il_0 j/n_0}.$$

Thus,

$$b_l/b_{l_0} = (l/l_0)^{(k/2)-1} \sum_{h=0}^{v} p^{h(2-k)} = \sum_{h=0}^{v} p^{h(k-2)}.$$

This relation between b_l and b_{l_0} can be summarized in the following identity, which is valid for any l_0 with $p|l_0, p^2\nmid l_0$:

$$\sum_{v=0}^{\infty} b_{l_0 p^{2v}} p^{-sv} = b_{l_0} \frac{1}{(1-p^{-s})(1-p^{k-2-s})}. \tag{2.24}$$

To see the equivalence of this identity to the formula derived above for b_l/b_{l_0}, it suffices to expand $(1-p^{-s})^{-1}$ and $(1-p^{k-2-s})^{-1}$ in geometric series and in that way conclude that the coefficient of p^{-vs} on the right in (2.24) is

$$b_{l_0} \sum_{h=0}^{v} p^{h(k-2)}.$$

Case (ii) $p\nmid l_0$. Again we fix n_0 with $p^2\nmid n_0$ and consider all terms S_n in the outer sum in (2.7) for which $n = n_0 p^{2h}$, $h = 0, 1, 2, \ldots$.

First suppose that $p\nmid n_0$. We take $j = j_0 p^{2h} + j_1 n_0$ as before, define $\delta(j)$ by (2.23), and find that

$$S_n = S_{n_0} p^{-hk} \sum_{0\le j_1<p^{2h}} \delta(j_1) e^{-2\pi ij_1 l_0 p^{2(v-h)}}.$$

As before, this sum can be rewritten in the form

$$\sum_{0\le j_1<p^{2h}} e^{-2\pi ij_1 p^{2(v-h)}} - \sum_{0\le j_2<p^{2h-1}} e^{-2\pi ij_2 p^{1+2(v-h)}},$$

which equals zero if $h > v$ and $p^{2h} - p^{2h-1}$ if $h \le v$. Thus,

$$\sum_{h=0}^{\infty} S_{n_0 p^{2h}} = S_{n_0}\left(1 + \sum_{h=1}^{v} p^{-hk}(p^{2h} - p^{2h-1})\right). \tag{2.25}$$

Now suppose that $p|n_0$, $n_0 = p\tilde{n}_0$. At this point we change our notation, replacing $\tilde{n}_0$ by n_0. That is, we shall determine $\Sigma_h S_{n_0 p^{2h+1}}$, where now n_0 is not divisible by p. As before, we set $j = j_0 p^{2h+1} + j_1 n_0$, and find that S_n, where $n = n_0 p^{2h+1}$, is equal to the product of

$$p^{-k/2}\varepsilon_{pn_0}^k n_0^{-k/2} \sum_{0\le j_0<n_0} \left(\frac{j_0 p}{n_0}\right) e^{-2\pi i j_0 l/n_0} \tag{2.26}$$

and

$$p^{-hk}\left(\frac{n_0}{p}\right) \sum_{0\le j_1<p^{2h+1}} \left(\frac{j_1}{p}\right) e^{-2\pi i j_1 l_0 p^{2(v-h)-1}}. \tag{2.27}$$

The latter sum is easily seen to vanish if either $h < v$ or $h > v$. So the only nonzero S_n occurs when $h = v$, i.e., $n = p^{2v+1}n_0$. Then the sum of (2.27) over $h = 0, 1, 2, \dots, v$ is equal to

$$p^{-vk}\left(\frac{-l_0}{p}\right)\left(\frac{n_0}{p}\right) \sum_{0\le j<p^{2v+1}} \left(\frac{j}{p}\right) e^{2\pi i j/p}$$

(after the change of variables $j = -l_0 j_1$); and this sum is p^{2v} times the Gauss sum $\varepsilon_p\sqrt{p}$. Meanwhile, (2.26) is equal to

$$p^{-k/2}(\varepsilon_{pn_0}/\varepsilon_{n_0})^k\left(\frac{p}{n_0}\right)S_{n_0}.$$

Hence, for $n = n_0 p^{2v+1}$ we have

$$\sum_{h=0}^{\infty} S_{n_0 p^{2h+1}} = p^{2v}\varepsilon_p\sqrt{p}\,p^{-vk}\left(\frac{-l_0}{p}\right)p^{-k/2}(\varepsilon_{pn_0}/\varepsilon_{n_0})^k\left(\frac{p}{n_0}\right)\left(\frac{n_0}{p}\right)S_{n_0}.$$

We check that $\varepsilon_p(\varepsilon_{pn_0}/\varepsilon_{n_0})^k(\frac{p}{n_0})(\frac{n_0}{p}) = (\frac{-1}{p})^{\lambda+1}$ by considering the four cases $p, n_0 \equiv \pm 1 \bmod 4$. Hence,

$$\sum_{h=0}^{\infty} S_{n_0 p^{2h+1}} = p^{v(2-k)-\lambda}\chi_{(-1)^\lambda l_0}(p)S_{n_0}.$$

Thus, combining this with (2.25), we find that the total contribution to the outer sum in (2.7) of all terms corresponding to n of the form n_0p^{2h} or n_0p^{2h+1} (for fixed n_0 not divisible by p) is equal to

$$S_{n_0}\left(1 + p^{v(2-k)-\lambda}\chi_{(-1)^\lambda l_0}(p) + \sum_{h=1}^{v} p^{-hk}(p^{2h} - p^{2h-1})\right). \tag{2.28}$$

In the case $l = l_0$, $v = 0$, this contribution is simply

$$S_{n_0}(1 + p^{-\lambda}\chi_{(-1)^\lambda l_0}(p)).$$

Let us temporarily abbreviate $\chi = \chi_{(-1)^\lambda l_0}(p)$. Then we compute that

$$\frac{b_l}{b_{l_0}} = \frac{p^{v(k-2)}\left(1 + p^{v(2-k)-\lambda}\chi + \sum_{h=1}^{v} p^{-hk}(p^{2h} - p^{2h-1})\right) \sum_{n_0>0\text{ odd}, p\nmid n_0} S_{n_0}}{(1 + p^{-\lambda}\chi) \sum_{n_0>0\text{ odd}, p\nmid n_0} S_{n_0}}$$

$$= \frac{\sum_{h=0}^{v} p^{h(k-2)} - \sum_{h=0}^{v-1} p^{h(k-2)-1} + p^{-\lambda}\chi}{1 + p^{-\lambda}\chi}.$$

By a simple algebraic computation, using $\chi^2 = 1$, we find that this last ratio is equal to

$$b_l/b_{l_0} = \sum_{h=0}^{v} p^{h(k-2)} - \chi p^{\lambda-1} \sum_{h=0}^{v-1} p^{h(k-2)}. \tag{2.29}$$

The relations (2.29) for $v = 0, 1, \ldots,$ can be expressed by the identity

$$\sum_{v=0}^{\infty} b_{l_0 p^{2v}} p^{-sv} = b_{l_0} \frac{1 - \chi_{(-1)^\lambda l_0}(p) p^{\lambda-1-s}}{(1-p^{-s})(1-p^{k-2-s})} \tag{2.30}$$

as we see in the usual manner by gathering coefficients of p^{-sv} on the right. Note that the relation (2.24) that we derived in case (i) is the same as (2.30), because $\chi_{(-1)^\lambda l_0}(p) = 0$ when $p | l_0$. Thus, in all cases we have (2.30).

The next proposition combines the identities (2.30) for all p odd and (2.22) for $p = 2$.

Proposition 8. *Let l_0 be a squarefree positive integer, let b_l denote the q-expansion coefficient of $F_{k/2}$, and set $\lambda = (k-1)/2$. Then*

$$\sum_{l_1=1}^{\infty} b_{l_0 l_1^2} l_1^{-s} = \frac{b_{l_0}}{1 - 2^{k-2-s}} \prod_{\text{odd } p} \frac{1 - \chi_{(-1)^\lambda l_0}(p) p^{\lambda-1-s}}{(1-p^{-s})(1-p^{k-2-s})}. \tag{2.31}$$

Proof. Let $l_1 = p_1^{v_1} \cdots p_r^{v_r}$. Let $f_p(s)$ denote the factor in the product corresponding to p (here $f_2(s) = (1 - 2^{k-2-s})^{-1}$). Then we must show that $b_{l_0 l_1^2}$ equals b_{l_0} times the coefficient of $l_1^{-s} = p_1^{-sv_1} \cdots p_r^{-sv_r}$ in $f_{p_1} \cdots f_{p_r}$. We use induction on r. For $r = 1$, this is precisely what (2.22) and (2.30) say. If $r > 1$ and $l_2 = p_1^{v_1} \cdots p_{r-1}^{v_{r-1}}$, we assume the result for $r - 1$, and then we have (by (2.22) or (2.30) with l_0 replaced by $l_0 l_2^2$):

$$\begin{aligned} b_{l_0 l_2^2 p_r^{2v_r}} &= b_{l_0 l_2^2} \cdot (\text{coefficient of } p_r^{-sv_r} \text{ in } f_{p_r}) \\ &= b_{l_0} \cdot (\text{coeff of } l_2^{-s} \text{ in } f_{p_1} \cdots f_{p_{r-1}})(\text{coeff of } p_r^{-sv_r} \text{ in } f_{p_r}) \end{aligned}$$

by the induction assumption. But the product of these two coefficients is the coefficient of l_1^{-s} in $f_{p_1} \cdots f_{p_r}$. This completes the proof. (Compare with the identity (5.8) for Hecke operators in §III.5.) □

In a similar manner, one can derive identities for the q-expansion coefficients a_l for $E_{k/2}$ as $l = l_0 l_1^2$ varies over integers with fixed squarefree part l_0. We shall only give the result. One again proceeds one prime at a time. For odd p, the same identity holds as for $F_{k/2}$:

$$\sum_{v=0}^{\infty} a_{l_0 p^{2v}} p^{-sv} = a_{l_0} \frac{1 - \chi_{(-1)^\lambda l_0}(p) p^{\lambda-1-s}}{(1-p^{-s})(1-p^{k-2-s})} \tag{2.32}$$

for fixed l_0, $p^2 \nmid l_0$. But for $p = 2$ the identity turns out to be a little more complicated than for $F_{k/2}$.

The most interesting Eisenstein series for $\Gamma_0(4)$ is the linear combination

$H_{k/2} = \zeta(1-2\lambda)(E_{k/2} + (1+i^k)2^{-k/2}F_{k/2})$ in Proposition 6. Let c_l be the l-th q-expansion coefficient of $H_{k/2}$. If $l = l_0 p^{2v}$, $p^2 \nmid l_0$, for an odd prime p, then, because of (2.30) and (2.32), we have the same identity for the c_l:

$$\sum_{v=0}^{\infty} c_{l_0 p^{2v}} p^{-sv} = c_{l_0} \frac{1 - \chi_{(-1)^\lambda l_0}(p) p^{\lambda-1-s}}{(1-p^{-s})(1-p^{k-2-s})}. \tag{2.33}$$

In the case $p = 2$, one has $b_{l_0 4^v} = 2^{v(k-2)} b_{l_0}$, and for $a_{l_0 4^v}$ one can derive formulas expressing $a_{l_0 4^v}$ in terms of $b_{l_0 4^v}$ (these formulas are given in [Cohen 1973]). Combining these formulas for the case $p = 2$, one obtains the following result: if l_0 is a positive integer such that either $(-1)^\lambda l_0 \equiv 1 \bmod 4$ or else $(-1)^\lambda l_0$ is four times an integer congruent to 2 or 3 mod 4, then the coefficients $c_{l_0 4^v}$, $v = 0, 1, 2, \ldots$, satisfy the same identity (2.33) with $p = 2$. Thus, under these circumstances the identity (2.33) holds for *all* primes p, including 2, in the case of $H_{k/2}$.

The identities (2.33) for all primes p lead to a relation similar to (2.31), where either l_0 is squarefree and $(-1)^\lambda l_0 \equiv 1 \bmod 4$, or else $l_0/4$ is squarefree and $(-1)^\lambda l_0/4 \equiv 2$ or $3 \bmod 4$. It can also be shown (see the problems below) that $c_l = 0$ for l which are not square multiples of such l_0. Since the coefficients c_{l_0} in these cases are precisely the values $L(\chi_{(-1)^\lambda l_0}, 1-\lambda)$, we obtain the following proposition.

Proposition 9. *The element $H_{k/2} \in M_{k/2}(\tilde{\Gamma}_0(4))$ given by $H_{k/2} = \zeta(1-2\lambda)$ $(E_{k/2} + (1+i^k)2^{-k/2}F_{k/2})$ has q-expansion coefficients c_l which can be determined by the identity*

$$\sum_{l_1=1}^{\infty} c_{l_0 l_1^2} l_1^{-s} = L(\chi_{(-1)^\lambda l_0}, 1-\lambda) \prod_{\text{all } p} \frac{1 - \chi_{(-1)^\lambda l_0}(p) p^{\lambda-1-s}}{(1-p^{-s})(1-p^{k-2-s})}, \tag{2.34}$$

where either l_0 is squarefree and $(-1)^\lambda l_0 \equiv 1 \bmod 4$, or else $l_0/4$ is squarefree and $(-1)^\lambda l_0/4 \equiv 2$ or $3 \bmod 4$. If l is not a square multiple of such an l_0, then $c_l = 0$. Finally, $c_0 = \zeta(1-2\lambda)$.

Thus, there are two very different elegant properties satisfied by the q-expansion coefficients of $H_{k/2}$. First, the coefficients corresponding to discriminants of quadratic fields are the values of the L-function for the corresponding quadratic character at the fixed negative integer $1 - \lambda = (3-k)/2$. Second, for a fixed discriminant, the coefficients c_l with l a square multiple of that discriminant satisfy an Euler product identity.

Both properties are closely analogous to the results about forms of half integral weight which we shall need later to describe Tunnell's work on the congruent number problem. The first property of $H_{k/2}$ is parallel to Waldspurger's theorem, asserting the existence of a modular form of half integral weight whose l_0-th q-expansion coefficients are closely related to $L(E_{l_0}, 1)$. The second property is an example of a very general correspon-

dence due to Shimura [1973a] between forms of half integral weight and forms of integral weight.

Let us look again at the Euler product (2.34) for $H_{k/2}$. The only part that depends on l_0 is the numerator $1 - \chi_{(-1)^\lambda l_0}(p)p^{\lambda-1-s}$. The remaining part of the Euler product

$$\prod_p (1 - p^{-s})^{-1}(1 - p^{k-2-s})^{-1},$$

depends only on $H_{k/2}$ and not on the choice of l_0. Observe that this is an Euler product we encountered before: it is the Euler product relating all the coefficients of the normalized Eisenstein series

$$-\frac{B_{k-1}}{2(k-1)}E_{k-1} = -\frac{B_{k-1}}{2(k-1)} + \sum_{n=1}^{\infty} \sigma_{k-2}(n)q^n$$

(see Problem 16 in §III.3). We say that $H_{k/2} \in M_{k/2}(\tilde{\Gamma}_0(4))$ corresponds to $(-B_{k-1}/(2(k-1)))E_{k-1} \in M_{k-1}(\Gamma)$ under the Shimura map.

We shall discuss the Shimura map in more detail later. The Shimura map applies to forms of half integer weight which have Euler products similar to (2.34). As in the case of forms of integral weight, Euler products arise from the property of being an eigenform for Hecke operators. But when we work with half integral weight, it turns out that we only have Hecke operators T_{n^2} whose index is a perfect square; all other T_n are identically zero. This is why, in the case of half integral weight, we get a weaker type of Euler product, which only connects coefficients whose indices differ by a perfect square multiple. Hecke operators on forms of half integral weight are the subject of the next section.

Problems

1. Verify the cusp condition for $E_{k/2}$, and find the value of $E_{k/2}$ and $F_{k/2}$ at all three cusps of $\Gamma_0(4)$.

2. Derive (2.16) from (2.15).

3. Derive (2.17).

4. Check that the constants α and β given in the proof of Proposition 6 are the solutions to the two equations giving $c_{|D|} = L(\chi_D, 1 - \lambda)$ in all cases.

5. Suppose that we looked for a linear combination $\alpha E_{k/2} + \beta F_{k/2}$ whose l-th q-expansion coefficient for l squarefree is equal to $L(\chi_{(-1)^\lambda l}, 1 - \lambda)$. (That is, when $(-1)^\lambda l \equiv 2$ or 3 mod 4, this L-function value is c_l rather than c_{4l}.) Show that no linear combination has this property for all such l.

6. Show that if $(-1)^\lambda l \equiv 2$ or 3 mod 4, then $c_l = 0$ in the q-expansion of $H_{k/2}$. (*Note.* If $M_{k/2}^+(\tilde{\Gamma}_0(4))$ denotes the subspace of $M_{k/2}(\tilde{\Gamma}_0(4))$ consisting of forms whose q-expansion coefficients satisfy this property, then it has been shown [Kohnen 1980] that the Shimura map gives an isomorphism of $M_{k/2}^+(\tilde{\Gamma}_0(4))$ with $M_{k-1}(\Gamma)$.)

7. What is the constant term in $H_{k/2}$? What is the coefficient of the first power of q in $H_{k/2}$? When is this latter coefficient zero?

8. Show that $H_{5/2} = \frac{1}{120}\Theta^5 - \frac{1}{6}\Theta F$, and show how this gives Proposition 7.

9. Show that for $k \geq 5$ odd and for suitable a and b, we have $\Theta^k - aE_{k/2} - bF_{k/2} \in S_{k/2}(\tilde{\Gamma}_0(4))$. Find a and b. Express Θ^5 and Θ^7 in terms of $E_{k/2}$ and $F_{k/2}$.

§3. Hecke operators on forms of half integer weight

We first recall how the Hecke operator T_n can be defined on forms of integral weight by means of double cosets. For simplicity, we review the case of the full modular group Γ.

For n a positive integer and $f \in M_k(\Gamma)$ we can define $T_n f$ as follows. Let Δ^n be the set of all 2×2—matrices with integer entries and determinant n. For any double coset $\Gamma\alpha\Gamma \subset \Delta^n$, where $\alpha \in \Delta^n$, we define $f|[\Gamma\alpha\Gamma]_k = \Sigma f|[\alpha\gamma_j]_k$, where the sum is over all right cosets $\Gamma\alpha\gamma_j \subset \Gamma\alpha\Gamma$; equivalently, γ_j runs through a complete set of right coset representatives of Γ modulo $\alpha^{-1}\Gamma\alpha \cap \Gamma$ (see Proposition 41 in §III.5). Then

$$T_n f \underset{\text{def}}{=} n^{(k/2)-1} \sum f|[\Gamma\alpha\Gamma]_k,$$

where the sum is over all double cosets of Γ in Δ^n.

There are not many double cosets of Γ in Δ^n; in fact, if n is squarefree there is only one. More precisely, we have the following proposition.

Proposition 10. *A complete set of double coset representatives of* Γ *in* Δ^n *is* $\{\left(\begin{smallmatrix} n_1 & 0 \\ 0 & n_1 n_0 \end{smallmatrix}\right)\}$, *where* n_0, n_1 *run through all positive integers such that* $n = n_0 n_1^2$. *In particular, if n is squarefree, then* $\Delta^n = \Gamma\left(\begin{smallmatrix} 1 & 0 \\ 0 & n \end{smallmatrix}\right)\Gamma$. *On the other hand, if* $n = p^2$ *is the square of a prime, then* $\Delta^{p^2} = \Gamma\left(\begin{smallmatrix} 1 & 0 \\ 0 & p^2 \end{smallmatrix}\right)\Gamma \cup p\Gamma$ (*where* $p\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} pa & pb \\ pc & pd \end{smallmatrix}\right)$).

PROOF. Consider the abelian group $\mathbb{Z}^2$ with standard basis generators $e_1 = \left(\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right)$, $e_2 = \left(\begin{smallmatrix} 0 \\ 1 \end{smallmatrix}\right)$. Any matrix $\alpha \in \Delta^n$ gives a subgroup of index n in $\mathbb{Z}^2$, denoted $\alpha\mathbb{Z}^2$, namely, the span of αe_1 and αe_2, the columns of α. Conversely, any subgroup of index n in $\mathbb{Z}^2$ can be obtained as $\alpha\mathbb{Z}^2$ for some α. Now fix any $\alpha \in \Delta^n$. By the elementary divisor theorem (see, e.g., [Van der Waerden 1970, Vol. 2, p. 4]), there exists a basis e_1', e_2' of $\mathbb{Z}^2$ such that the subgroup $\alpha\mathbb{Z}^2$ is the span of $n_1 e_1'$ and $n_1 n_0 e_2'$ for some positive integers n_0, n_1 with $n = n_0 n_1^2$. Let $\gamma_1 \in \Gamma$ be the change of basis matrix from e_1, e_2 to e_1', e_2', and let $\gamma_2^{-1} \in \Gamma$ be the change of basis matrix from αe_1, αe_2 to $n_1 e_1'$, $n_1 n_0 e_2'$. Thus, $\alpha\gamma_2^{-1} = [n_1 e_1', n_1 n_0 e_2'] = \gamma_1\left(\begin{smallmatrix} n_1 & 0 \\ 0 & n_1 n_0 \end{smallmatrix}\right)$, i.e., $\alpha = \gamma_1\left(\begin{smallmatrix} n_1 & 0 \\ 0 & n_1 n_0 \end{smallmatrix}\right)\gamma_2 \in \Gamma\left(\begin{smallmatrix} n_1 & 0 \\ 0 & n_1 n_0 \end{smallmatrix}\right)\Gamma$. Since $\alpha \in \Delta^n$ is arbitrary, this proves that the indicated double cosets exhaust Δ^n. Conversely, it is easy to see that these double cosets are disjoint: in fact, $\alpha \in \Gamma\left(\begin{smallmatrix} n_1 & 0 \\ 0 & n_1 n_0 \end{smallmatrix}\right)\Gamma = n_1\Gamma\left(\begin{smallmatrix} 1 & 0 \\ 0 & n_0 \end{smallmatrix}\right)\Gamma$ if and only if n_1 is the greatest common divisor

of all entries in the matrix α. This is because, if $\alpha = \gamma_1\alpha'\gamma_2$ with $\gamma_1, \gamma_2 \in \Gamma$, then $d|$all entries of α' implies $d|$all entries of α; conversely, because $\alpha' = \gamma_1^{-1}\alpha\gamma_2^{-1}$, it follows that $d|$all entries of α implies $d|$all entries of α'. This completes the proof. □

In §III.5 we discussed the Hecke operators on forms of integral weight for the congruence subgroup $\Gamma_1(N)$. In that case one lets $\Delta^n = \Delta^n(N, \{1\}, \mathbb{Z})$ (see (5.23) in §III.5), i.e., Δ^n is the set of matrices of determinant n which are congruent to $\left(\begin{smallmatrix}1 & *\\ 0 & n\end{smallmatrix}\right)$ modulo N. For $f \in M_k(\Gamma_1(N))$ one defines

$$T_n f \underset{\text{def}}{=} n^{(k/2)-1} \sum f|[\Gamma_1(N)\alpha\Gamma_1(N)]_k,$$

where the sum is over all double cosets of $\Gamma_1(N)$ in Δ^n.

Proposition 11. *If* g.c.d.$(n, N) = 1$, *then a complete set of double coset representatives of* $\Gamma_1(N)$ *in* $\Delta^n = \Delta^n(N, \{1\}, \mathbb{Z})$ *is* $\{\sigma_{n_1}\left(\begin{smallmatrix}n_1 & 0\\ 0 & n_1n_0\end{smallmatrix}\right)\}$, *where* n_0, n_1 *are as in Proposition* 10 *and* σ_{n_1} *is a fixed element of* Γ *such that* $\sigma_{n_1} \equiv \left(\begin{smallmatrix}1/n_1 & 0\\ 0 & n_1\end{smallmatrix}\right)$ *modulo* N.

Proof. If $\alpha \in \Delta^n$, we know by Proposition 10 that $\alpha = n_1\gamma_1\left(\begin{smallmatrix}1 & 0\\ 0 & n_0\end{smallmatrix}\right)\gamma_2$, where $\gamma_1, \gamma_2 \in \Gamma$ and n_1 is the greatest common divisor of the entries of α. We must show that α can be written in the form $n_1\gamma_1''\sigma_{n_1}\left(\begin{smallmatrix}1 & 0\\ 0 & n_0\end{smallmatrix}\right)\gamma_2'$ with $\gamma_1'', \gamma_2' \in \Gamma_1(N)$, or equivalently, in the form $n_1\sigma_{n_1}\gamma_1'\left(\begin{smallmatrix}1 & 0\\ 0 & n_0\end{smallmatrix}\right)\gamma_2'$ with $\gamma_1' = \sigma_{n_1}^{-1}\gamma_1''\sigma_{n_1} \in \Gamma_1(N)$. We suppose that $n_1 = 1$, i.e., $n = n_0$; the general case is completely similar.

Lemma. *If* g.c.d.$(n, N) = 1$, *then a set* $\{\tau_j\}$ *of right coset representatives for* Γ modulo $\Gamma_1(N)$ *can be chosen so that* $\tau_j \in \Gamma_0(n)$.

Proof of Lemma. Let $\tau = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) \in \Gamma$ be any coset representative. We look for $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) \in \Gamma_1(N)$ such that $\tau' = \left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)\tau \in \Gamma_0(n)$, in which case we merely replace τ by τ'. It is easy to see that there exist u and v relatively prime such that n divides $au + cv$. Let $\gamma = u + (j_1 + j_2N)n$, $\delta = v + ln$, where j_1 and l are chosen so that $u + j_1 n \equiv 0 \bmod N$, $v + ln \equiv 1 \bmod N$ (this is possible because g.c.d.$(n, N) = 1$). If we show that j_2 can be chosen so that γ and δ are relatively prime, then we can find α, β such that $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right) \in \Gamma_1(N)$ and $\left(\begin{smallmatrix}\alpha & \beta\\ \gamma & \delta\end{smallmatrix}\right)\tau \in \Gamma_0(n)$, because $\gamma a + \delta c \equiv au + cv \equiv 0 \bmod n$. But if $u + j_1 n + j_2 Nn$ and $v + ln$ have a common factor, we first note that such a divisor must be prime to N (since $v + ln \equiv 1 \bmod N$) and also prime to n (since g.c.d.$(u, v) = 1$). Then if P is the product of all prime divisors of $v + ln$ not dividing Nn, we can find j_2 such that $u + j_1 n + j_2 Nn \equiv 1 \bmod P$. In this way we can find the required j_2.

We now return to the proof of the proposition. We have $\alpha = \gamma_1\left(\begin{smallmatrix}1 & 0\\ 0 & n\end{smallmatrix}\right)\gamma_2$ with $\gamma_1, \gamma_2 \in \Gamma$. Using the lemma, we write $\gamma_1 \in \Gamma$ in the form $\gamma_1 = \gamma_1'\gamma_1''$ where $\gamma_1'' \in \Gamma_0(n)$ and $\gamma_1' \in \Gamma_1(N)$. Since $\gamma_1'' \in \Gamma_0(n)$, it follows that $\left(\begin{smallmatrix}1 & 0\\ 0 & n\end{smallmatrix}\right)^{-1}\gamma_1''\left(\begin{smallmatrix}1 & 0\\ 0 & n\end{smallmatrix}\right) \in \Gamma$. We write

$$\alpha = \gamma_1' \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix}^{-1} \gamma_1'' \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \gamma_2 = \gamma_1' \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \gamma_2',$$

where $\gamma_2' = \left(\begin{smallmatrix} 1 & 0 \\ 0 & n \end{smallmatrix}\right)^{-1} \gamma_1'' \left(\begin{smallmatrix} 1 & 0 \\ 0 & n \end{smallmatrix}\right) \gamma_2 \in \Gamma$. It remains to show that $\gamma_2' \in \Gamma_1(N)$. But since $\alpha \in \Delta^n$ is congruent to $\left(\begin{smallmatrix} 1 & * \\ 0 & n \end{smallmatrix}\right)$ modulo N, we have modulo N

$$\begin{pmatrix} 1 & * \\ 0 & n \end{pmatrix} \equiv \begin{pmatrix} 1 & * \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix} \gamma_2' \equiv \begin{pmatrix} 1 & * \\ 0 & n \end{pmatrix} \gamma_2',$$

from which it immediately follows that $\gamma_2' \in \Gamma_1(N)$. This completes the proof. □

We now look at the analogous construction for forms of half integer weight. Recall that in that case we must work with the groups $\tilde{\Gamma}_0(4) \subset G^1 \subset G$, where

$$G = \left\{ (\alpha, \phi(z)) \,|\, \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in GL_2^+(\mathbb{Q}), \right.$$

$$\left. \phi(z)^2 = t(cz + d)/\sqrt{\det \alpha} \text{ for some } t = \pm 1 \right\};$$

$$G^1 = \{(\alpha, \phi(z)) \in G \,|\, \alpha \in \Gamma\};$$

$$\tilde{\Gamma}_0(4) = \left\{ (\alpha, j(\alpha, z)) \,|\, \alpha = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(4), j(\alpha, z) = \left(\frac{c}{d}\right) \varepsilon_d^{-1} \sqrt{cz + d} \right\}.$$

Suppose that $4|N$ and $f \in M_{k/2}(\tilde{\Gamma}_1(N))$. Let n be any positive integer prime to N. In the case of integer weight we defined T_n by considering double cosets of the form $\Gamma_1(N)\sigma_{n_1}\left(\begin{smallmatrix} 1 & 0 \\ 0 & n_0 \end{smallmatrix}\right)\Gamma_1(N)$, where $n = n_0 n_1^2$. So for half integer weight one considers double cosets of the form $\tilde{\Gamma}_1(N)\tilde{\sigma}_{n_1}\xi_{n_0}\tilde{\Gamma}_1(N)$, where $\xi_{n_0} \in G$ is any lifting of $\left(\begin{smallmatrix} 1 & 0 \\ 0 & n_0 \end{smallmatrix}\right)$, i.e., $\xi_{n_0} = (\left(\begin{smallmatrix} 1 & 0 \\ 0 & n_0 \end{smallmatrix}\right), t n_0^{1/4})$ for some $t = \pm 1, \pm i$. Since t would turn out only to affect our definition by the constant multiple t^k, for simplicity we agree always to take $t = 1$. Also for simplicity we consider the case $n_1 = 1$, $n_0 = n$; the general case is completely similar. So we now examine the action of the double coset $\tilde{\Gamma}_1(N)\xi_n\tilde{\Gamma}_1(N)$ on f, where

$$\xi_n = \left(\begin{pmatrix} 1 & 0 \\ 0 & n \end{pmatrix}, \sqrt[4]{n} \right).$$

That is, we compute

$$f|[\tilde{\Gamma}_1(N)\xi_n\tilde{\Gamma}_1(N)]_{k/2} \underset{\text{def}}{=} \sum_j f|[\xi_n\tilde{\gamma}_j]_{k/2}, \tag{3.1}$$

where the sum is over all distinct right cosets of $\tilde{\Gamma}_1(N)$ in our double coset, i.e., over a set $\{\tilde{\gamma}_j\}$ of right coset representatives of $\tilde{\Gamma}_1(N)$ modulo $\tilde{\Gamma}'' = \xi_n^{-1}\tilde{\Gamma}_1(N)\xi_n \cap \tilde{\Gamma}_1(N)$ (see Proposition 41 in §III.5).

Proposition 12. *If n is a positive integer prime to N which is not a perfect square, then* $f|[\tilde{\Gamma}_1(N)\xi_n\tilde{\Gamma}_1(N)]_{k/2} = 0$.

PROOF. Given $\alpha \in GL_2^+(\mathbb{Q})$ and $\xi = (\alpha, \phi) \in G$, we construct a map from $\Gamma' \underset{\text{def}}{=} \alpha^{-1}\Gamma_1(N)\alpha \cap \Gamma_1(N)$ to $T = \{\pm 1, \pm i\}$ in the following way. Given $\gamma = \alpha^{-1}\gamma_1\alpha$ with $\gamma, \gamma_1 \in \Gamma_1(N)$, observe that $\tilde{\gamma}$ and $\xi^{-1}\tilde{\gamma}_1\xi \in G_1$ both have the same projection in Γ, and so they differ by an element of the form $((\begin{smallmatrix}1 & 0\\ 0 & 1\end{smallmatrix}), t)$, i.e.,

$$\xi^{-1}\tilde{\gamma}_1\xi = \tilde{\gamma}(1, t).$$

For fixed α and ξ, we consider the map that associates to γ the number t. One verifies (see the exercises below) that $t(\gamma)$ is a group homomorphism from Γ' to T, that it depends only on α and not on the choice of $\phi(z)$ in $\xi = (\alpha, \phi)$, and that in the case $\alpha = (\begin{smallmatrix}1 & 0\\ 0 & n\end{smallmatrix})$ we have

$$t(\gamma) = \left(\frac{d}{n}\right) \quad \text{for} \quad \gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma' = \alpha^{-1}\Gamma_1(N)\alpha \cap \Gamma_1(N). \tag{3.2}$$

Let $K \subset \Gamma'$ denote the kernel of this map t.

Recall the definition $\tilde{\Gamma}'' = \xi_n^{-1}\tilde{\Gamma}_1(N)\xi_n \cap \tilde{\Gamma}_1(N)$. We claim that $\tilde{K} = \tilde{\Gamma}''$, i.e., if $\tilde{\gamma} \in \tilde{\Gamma}_1(N)$ is of the form $\xi_n^{-1}\tilde{\gamma}_1\xi_n$ with $\gamma_1 \in \Gamma_1(N)$, then $\gamma \in K$, and conversely. To see this, first suppose that $\gamma, \gamma_1 \in \Gamma_1(N)$ and $\tilde{\gamma} = \xi_n^{-1}\tilde{\gamma}_1\xi_n$. Applying the projection $P\colon G \to GL_2^+(\mathbb{Q})$ gives $\gamma = (\begin{smallmatrix}1 & 0\\ 0 & n\end{smallmatrix})^{-1}\gamma_1(\begin{smallmatrix}1 & 0\\ 0 & n\end{smallmatrix})$, i.e., $\gamma \in \Gamma' = \alpha^{-1}\Gamma_1(N)\alpha \cap \Gamma_1(N)$. Since $\xi_n^{-1}\tilde{\gamma}_1\xi_n = \tilde{\gamma} = \tilde{\gamma}(1, t)$, it follows that $\gamma \in \operatorname{Ker} t$. Conversely, if $\gamma \in K \subset \Gamma'$, so that $\gamma = \alpha^{-1}\gamma_1\alpha$ and $\xi_n^{-1}\tilde{\gamma}_1\xi_n = \tilde{\gamma}(1, t)$ with $t = 1$, it immediately follows that $\tilde{\gamma} \in \tilde{\Gamma}''$.

Thus, in general, $\tilde{\Gamma}'' = \tilde{K}$ is *smaller* than $\tilde{\Gamma}'$, i.e., the intersection of $\xi_n^{-1}\tilde{\Gamma}_1(N)\xi_n$ with $\tilde{\Gamma}_1(N)$ is a subgroup of the lifting of $\Gamma' = \alpha^{-1}\Gamma_1(N)\alpha \cap \Gamma_1(N)$. This subgroup $\tilde{\Gamma}''$ is all of $\tilde{\Gamma}'$ if and only if the map t is trivial. In our case $t(\gamma) = (\frac{d}{n})$, so that t is trivial if and only if n is a perfect square. (We are always assuming here that n is prime to N; the case $n = p|N$ is treated in Problem 3 below.) If n is not a perfect square, then $\tilde{\Gamma}''$ is a subgroup of index 2 in $\tilde{\Gamma}'$. In that case let $\tilde{\Gamma}' = \tilde{\Gamma}'' \cup \tilde{\Gamma}''\tilde{\tau}$ be a right coset decomposition; thus, $\tau = \alpha^{-1}\tau_1\alpha$ and $\tilde{\tau} = \xi_n^{-1}\tilde{\tau}_1\xi_n \cdot (1, -1)$. Let $\Gamma_1(N) = \bigcup_j \Gamma'\gamma_j$ be a right coset decomposition of $\Gamma_1(N)$ modulo Γ'. Then

$$\tilde{\Gamma}_1(N) = \bigcup_j \tilde{\Gamma}''\tilde{\gamma}_j \cup \bigcup_j \tilde{\Gamma}''\tilde{\tau}\tilde{\gamma}_j$$

is a right coset decomposition of $\tilde{\Gamma}_1(N)$ modulo $\tilde{\Gamma}'' = \xi_n^{-1}\tilde{\Gamma}_1(N)\xi_n \cap \tilde{\Gamma}_1(N)$. By the definition and Proposition 41 of §III.3, we have

$$f|[\tilde{\Gamma}_1(N)\xi_n\tilde{\Gamma}_1(N)]_{k/2} = \sum_j f|[\xi_n\tilde{\gamma}_j]_{k/2} + \sum_j f|[\xi_n\tilde{\tau}\tilde{\gamma}_j]_{k/2}.$$

But for each j we have

$$\begin{aligned} f|[\xi_n\tilde{\tau}\tilde{\gamma}_j]_{k/2} &= f|[\xi_n\tilde{\tau}\xi_n^{-1}\xi_n\tilde{\gamma}_j]_{k/2} \\ &= f|[\tilde{\tau}_1(1, -1)\xi_n\tilde{\gamma}_j]_{k/2} \\ &= f|[(1, -1)\xi_n\tilde{\gamma}_j]_{k/2}, \end{aligned}$$

because f is invariant under $[\tilde{\tau}_1]_{k/2}$ for $\tilde{\tau}_1 \in \tilde{\Gamma}_1(N)$. Since $[(1, -1)]_{k/2} =$

$(-1)^k = -1$ by definition, we have

$$f|[\xi_n\tilde{\gamma}_j]_{k/2} + f|[\xi_n\tilde{\tau}\tilde{\gamma}_j]_{k/2} = f|[\xi_n\tilde{\gamma}_j]_{k/2} - f|[\xi_n\tilde{\gamma}_j]_{k/2} = 0.$$

This completes the proof of Proposition 12. □

The same argument will show that $f|[\tilde{\Gamma}_1(N)\tilde{\sigma}_{n_1}\xi_{n_0}\tilde{\Gamma}_1(N)]_{k/2} = 0$ if n_0 is a positive nonsquare integer prime to N.

Because of Proposition 12, we can only work with a Hecke operator on $M_{k/2}(\tilde{\Gamma}_1(N))$ of index prime to N if that index is a perfect square n^2; otherwise, the Hecke operator is identically zero. Recall that the building blocks for the Hecke operators in the integer weight case are the T_p for p a prime (see Proposition 32 of §III.5). Similarly, in the half integer weight case, the building blocks for the T_{n^2}, g.c.d.$(n, N) = 1$, are the T_{p^2}, $p\nmid N$.

So let us examine in detail the action of T_{p^2} on $M_{k/2}(\tilde{\Gamma}_1(N))$, where T_{p^2} is defined on $f\in M_{k/2}(\tilde{\Gamma}_1(N))$ in the following way:

$$T_{p^2}f \underset{\text{def}}{=} p^{(k/2)-2}f|[\tilde{\Gamma}_1(N)\xi_{p^2}\tilde{\Gamma}_1(N)]_{k/2}, \quad \text{where} \quad \xi_{p^2} = \left(\begin{pmatrix}1 & 0\\ 0 & p^2\end{pmatrix}, \sqrt{p}\right). \tag{3.3}$$

This definition is not quite the immediate analog of T_{p^2} acting on $f\in M_k(\Gamma_1(N))$, which is given by $p^{k-2}(f|[\Gamma_1(N)(\begin{smallmatrix}1 & 0\\ 0 & p^2\end{smallmatrix})\Gamma_1(N)]_k + \chi(p)f)$ for $f\in M_k(N, \chi)$, because $\Delta^{p^2}(N, \{1\}, \mathbb{Z}) = \Gamma_1(N)(\begin{smallmatrix}1 & 0\\ 0 & p^2\end{smallmatrix})\Gamma_1(N) \cup \Gamma_1(N)\sigma_p(\begin{smallmatrix}p & 0\\ 0 & p\end{smallmatrix})\cdot\Gamma_1(N)$ by Proposition 11, and $f|[\sigma_p(\begin{smallmatrix}p & 0\\ 0 & p\end{smallmatrix})\Gamma_1(N)]_k = \chi(p)f$. Besides replacing k by $k/2$ and $\Gamma_1(N)$ and $(\begin{smallmatrix}1 & 0\\ 0 & p^2\end{smallmatrix})$ by their liftings $\tilde{\Gamma}_1(N)$ and ξ_{p^2}, we also drop the trivial double coset $\Gamma_1(N)\sigma_p(\begin{smallmatrix}p & 0\\ 0 & p\end{smallmatrix})\Gamma_1(N)$. In the case of integer weight, when T_p and not T_{p^2} is the building block, T_{p^2} involves two double cosets; but for half integer weight we shall agree to take only the nontrivial double coset.

As in the case of integer weight, in studying $M_{k/2}(\tilde{\Gamma}_1(N))$ it is convenient to decompose it into χ-components. If χ is any Dirichlet character modulo N, recall from §1 that

$$M_{k/2}(\tilde{\Gamma}_0(N), \chi) \underset{\text{def}}{=} \left\{ f\in M_{k/2}(\tilde{\Gamma}_1(N)) \mid f|[\tilde{\gamma}]_{k/2} = \chi(d)f \quad \text{for all} \quad \gamma = \begin{pmatrix}a & b\\ c & d\end{pmatrix}\in\Gamma_0(N) \right\};$$

and that

$$M_{k/2}(\tilde{\Gamma}_1(N)) = \oplus_\chi M_{k/2}(\tilde{\Gamma}_0(N), \chi),$$

where the sum is over all even Dirichlet characters modulo N (of course, $M_{k/2}(\tilde{\Gamma}_0(N), \chi) = 0$ if χ is an odd character, since $f = f|[\widetilde{-1}]_{k/2} = \chi(-1)f$).

By an argument which is completely analogous to that in the integer weight case, one can show that T_{p^2} takes $M_{k/2}(\tilde{\Gamma}_1(N))$ to itself; moreover (see Problem 5 below), it preserves the χ-component.

We now fix a Dirichlet character χ modulo N, suppose that $f\in M_{k/2}(\tilde{\Gamma}_0(N),$

χ), and evaluate (3.3) explicitly. Our purpose is to express the Fourier coefficients for $T_{p^2}f$ in terms of those for f.

By the lemma in the proof of Proposition 43 in §III.5, we know that $\Delta^{p^2}(N, \{1\}, \mathbb{Z})$ is the disjoint union of the right cosets of $\Gamma_1(N)$ with representatives

$$\alpha_b = \begin{pmatrix} 1 & b \\ 0 & p^2 \end{pmatrix} \quad \text{for} \quad 0 \le b < p^2;$$

$$\beta_h = \sigma_p \begin{pmatrix} p & h \\ 0 & p \end{pmatrix} \quad \text{for} \quad 0 \le h < p; \qquad \text{and} \quad \tau = \sigma_{p^2} \begin{pmatrix} p^2 & 0 \\ 0 & 1 \end{pmatrix}.$$

Here for n prime to N, σ_n denotes a fixed element of Γ such that $\sigma_n \equiv \left(\begin{smallmatrix} 1/n & 0 \\ 0 & n \end{smallmatrix}\right)$ mod N. The trivial double coset is the right coset corresponding to β_0; thus, we have the disjoint union

$$\Gamma_1(N)\begin{pmatrix} 1 & 0 \\ 0 & p^2 \end{pmatrix}\Gamma_1(N) = \bigcup_{b=0}^{p^2-1} \Gamma_1(N)\alpha_b \cup \bigcup_{h=1}^{p-1} \Gamma_1(N)\beta_h \cup \Gamma_1(N)\tau.$$

In order to evaluate (3.3) using the definition (3.1), we need the right coset decomposition of $\tilde{\Gamma}_1(N)\xi_{p^2}\tilde{\Gamma}_1(N)$ in the form $\bigcup_j \tilde{\Gamma}_1(N)\xi_{p^2}\tilde{\gamma}_j$. If we could write each α_b, β_h, τ in the form $\left(\begin{smallmatrix} 1 & 0 \\ 0 & p^2 \end{smallmatrix}\right)\gamma$, where $\gamma \in \Gamma_1(N)$, then the γ would be right coset representatives of $\Gamma_1(N)$ modulo $\left(\begin{smallmatrix} 1 & 0 \\ 0 & p^2 \end{smallmatrix}\right)^{-1}\Gamma_1(N)\left(\begin{smallmatrix} 1 & 0 \\ 0 & p^2 \end{smallmatrix}\right) \cap \Gamma_1(N)$, and so the corresponding liftings $\tilde{\gamma}$ would be right coset representatives of $\tilde{\Gamma}_1(N)$ modulo $\xi_{p^2}^{-1}\tilde{\Gamma}_1(N)\xi_{p^2} \cap \tilde{\Gamma}_1(N)$ (see Problem 1(b) below). In that case, (3.3) is equal to $\Sigma_\gamma f|[\xi_{p^2}\tilde{\gamma}]_{k/2}$.

Here we may adjust α_b, β_h, or τ by multiplying on the left by any $\gamma'^{-1} \in \Gamma_1(N)$—this merely replaces one right coset representative of $\Gamma_1(N)$ in $\Gamma_1(N)\left(\begin{smallmatrix} 1 & 0 \\ 0 & p^2 \end{smallmatrix}\right)\Gamma_1(N)$ by another representative of the same coset. In other words, it suffices to write α_b, β_h, τ in the form $\gamma'\left(\begin{smallmatrix} 1 & 0 \\ 0 & p^2 \end{smallmatrix}\right)\gamma$, where $\gamma, \gamma' \in \Gamma_1(N)$, and then compute

$$T_{p^2}f(z) = p^{(k/2)-2} \sum_\gamma f(z)|[\xi_{p^2}\tilde{\gamma}]_{k/2}. \tag{3.4}$$

The result will be the following proposition. We shall go through the detailed computation of $T_{p^2}f$ after the statement of the proposition.

Proposition 13. *Suppose that* $4|N$, χ *is a Dirichlet character modulo* N, $p \nmid N$ *is a prime, and* $k = 2\lambda + 1$ *is a positive odd integer. Let* $f(z) = \Sigma_{n=0}^\infty a_n e^{2\pi i n z} \in M_{k/2}(\tilde{\Gamma}_0(N), \chi)$. *Then*

$$T_{p^2}f(z) = \sum_{n=0}^\infty b_n e^{2\pi i n z},$$

where

$$b_n = a_{p^2 n} + \chi(p)\left(\frac{(-1)^\lambda n}{p}\right)p^{\lambda-1}a_n + \chi(p^2)p^{k-2}a_{n/p^2} \tag{3.5}$$

(*here we take* $a_{n/p^2} = 0$ *if* $p^2 \nmid n$).

PROOF. As explained above, we must write each α_b, β_h, τ in the form $\gamma'\left(\begin{smallmatrix}1 & 0\\ 0 & p^2\end{smallmatrix}\right)\gamma$ with $\gamma, \gamma' \in \Gamma_1(N)$. First, $\alpha_b = \left(\begin{smallmatrix}1 & b\\ 0 & p^2\end{smallmatrix}\right) = \left(\begin{smallmatrix}1 & 0\\ 0 & p^2\end{smallmatrix}\right)\left(\begin{smallmatrix}1 & b\\ 0 & 1\end{smallmatrix}\right)$, i.e., for α_b we can take $\gamma' = 1$, $\gamma = \left(\begin{smallmatrix}1 & b\\ 0 & 1\end{smallmatrix}\right)$. Thus, the contribution to (3.4) from all of the α_b is equal to

$$\begin{aligned} &p^{(k/2)-2}\sum_{b=0}^{p^2-1} f(z)\left|\left[\left(\begin{pmatrix}1 & 0\\ 0 & p^2\end{pmatrix}, \sqrt{p}\right)\left(\begin{pmatrix}1 & b\\ 0 & 1\end{pmatrix}, 1\right)\right]_{k/2}\right. \\ &= p^{(k/2)-2}\sum_{b=0}^{p^2-1} f\left(\frac{z+b}{p^2}\right)(\sqrt{p})^{-k} \\ &= \frac{1}{p^2}\sum_{b=0}^{p^2-1} f\left(\frac{z+b}{p^2}\right). \end{aligned}$$

Since

$$\sum_{b=0}^{p^2-1} e^{2\pi i n(z+b)/p^2} = \begin{cases} 0 & \text{if } p^2 \nmid n; \\ p^2 e^{2\pi i n z/p^2} & \text{if } p^2 | n, \end{cases}$$

we see that this contribution is $\Sigma_n a_{p^2n} e^{2\pi i n z}$. Thus, the α_b give the first term on the right in (3.5).

We next evaluate the contribution of τ. Since g.c.d.$(p^2, N) = 1$, we can find two integers u, v such that $up^2 + vN = 1$. Then we have

$$\tau = \sigma_{p^2}\begin{pmatrix}p^2 & 0\\ 0 & 1\end{pmatrix} = \sigma_{p^2}\begin{pmatrix}p^2 & -v\\ N & u\end{pmatrix}\begin{pmatrix}1 & 0\\ 0 & p^2\end{pmatrix}\begin{pmatrix}p^2u & v\\ -N & 1\end{pmatrix},$$

i.e., we take $\gamma' = \sigma_{p^2}\gamma''$ with $\gamma'' = \left(\begin{smallmatrix}p^2 & -v\\ N & u\end{smallmatrix}\right)$, and $\gamma = \left(\begin{smallmatrix}p^2u & v\\ -N & 1\end{smallmatrix}\right)$; then $\gamma', \gamma \in \Gamma_1(N)$. We have

$$\begin{aligned} f(z)|[\xi_{p^2}\tilde{\gamma}]_{k/2} &= f(z)|[\tilde{\sigma}_{p^2}\tilde{\gamma}''\xi_{p^2}\tilde{\gamma}]_{k/2} \\ &= \chi(p^2)f(z)\left|\left[(\gamma'', j(\gamma'', z))\cdot\left(\begin{pmatrix}1 & 0\\ 0 & p^2\end{pmatrix}, \sqrt{p}\right)\cdot(\gamma, j(\gamma, z))\right]_{k/2}\right.. \end{aligned}$$

A simple computation of the product inside $[\ \ ]_{k/2}$ gives $(\left(\begin{smallmatrix}p^2 & 0\\ 0 & 1\end{smallmatrix}\right), p^{-1/2})$. Thus, the contribution to (3.4) from τ is equal to

$$\begin{aligned} p^{(k/2)-2}\chi(p^2)f(z)\left|\left[\begin{pmatrix}p^2 & 0\\ 0 & 1\end{pmatrix}, p^{-1/2}\right]_{k/2}\right. &= p^{(k/2)-2}\chi(p^2)p^{k/2}f(p^2z) \\ &= p^{k-2}\chi(p^2)\sum_n a_n e^{2\pi i n p^2 z}. \end{aligned}$$

Hence, τ gives the third term on the right in (3.5).

Finally, we evaluate the contribution of the β_h, $0 < h < p$. Again we want to write $\beta_h = \sigma_p\left(\begin{smallmatrix}p & h\\ 0 & p\end{smallmatrix}\right)$ in the form $\sigma_p\gamma''\left(\begin{smallmatrix}1 & 0\\ 0 & p^2\end{smallmatrix}\right)\gamma$ with γ and $\sigma_p\gamma''$ in $\Gamma_1(N)$. So we look for $\gamma = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) \in \Gamma_1(N)$ and $\gamma'' \in \Gamma_0(N)$ such that $\left(\begin{smallmatrix}p & h\\ 0 & p\end{smallmatrix}\right) = \gamma''\left(\begin{smallmatrix}1 & 0\\ 0 & p^2\end{smallmatrix}\right)\cdot\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right)$, i.e.,

$$\begin{pmatrix}p & h\\ 0 & p\end{pmatrix}\begin{pmatrix}a & b\\ c & d\end{pmatrix}^{-1}\begin{pmatrix}1 & 0\\ 0 & p^2\end{pmatrix}^{-1} = \begin{pmatrix}p & h\\ 0 & p\end{pmatrix}\begin{pmatrix}d & -b/p^2\\ -c & a/p^2\end{pmatrix} \in \Gamma_0(N).$$

Clearly $p|a$, so we write $a = pa'$, $c = Nc'$. Thus, we need $-b/p + ha'/p \in \mathbb{Z}$.

So first choose any integer a' prime to p such that $a = pa' \equiv 1 \pmod N$. Then any $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma$ with $a = pa'$, $c = Nc'$ will be in $\Gamma_1(N)$. Next choose any b prime to a' such that $b \equiv ha' \pmod p$ (this is clearly possible). Since g.c.d.$(Nb, pa') = 1$, we can find c' and d such that $dpa' - c'Nb = 1$. Thus, $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) = \left(\begin{smallmatrix} pa' & b \\ Nc' & d \end{smallmatrix}\right) \in \Gamma_1(N)$, and

$$\begin{pmatrix} p & h \\ 0 & p \end{pmatrix} = \begin{pmatrix} pd - hc & (a'h - b)/p \\ -pc & a' \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & p^2 \end{pmatrix} \begin{pmatrix} pa' & b \\ c & d \end{pmatrix},$$

which we denote $\gamma''\left(\begin{smallmatrix} 1 & 0 \\ 0 & p^2 \end{smallmatrix}\right)\gamma$. As in the previous paragraph, where we evaluated the contribution of τ, we obtain

$$f(z)|[\xi_{p^2}\tilde{\gamma}]_{k/2} = \chi(p) f(z) \left| \left[(\gamma'', j(\gamma'', z)) \cdot \left(\begin{pmatrix} 1 & 0 \\ 0 & p^2 \end{pmatrix}, \sqrt{p} \right) \cdot (\gamma, j(\gamma, z)) \right]_{k/2} \right. .$$

We compute

$$\left(\begin{pmatrix} pd - hc & (a'h - b)/p \\ -pc & a' \end{pmatrix}, \left(\frac{-pc}{a'}\right) \varepsilon_{a'}^{-1} \sqrt{-pcz + a'} \right)$$

$$\cdot \left(\begin{pmatrix} 1 & 0 \\ 0 & p^2 \end{pmatrix}, \sqrt{p} \right) \cdot \left(\begin{pmatrix} pa' & b \\ c & d \end{pmatrix}, \left(\frac{c}{d}\right) \varepsilon_d^{-1} \sqrt{cz + d} \right)$$

$$= \left(\begin{pmatrix} pd - hc & (a'h - b)p \\ -pc & a'p^2 \end{pmatrix}, \left(\frac{-pc}{a'}\right) \varepsilon_{a'}^{-1} \sqrt{-cz + a} \right)$$

$$\cdot \left(\begin{pmatrix} pa' & b \\ c & d \end{pmatrix}, \left(\frac{c}{d}\right) \varepsilon_d^{-1} \sqrt{cz + d} \right)$$

$$= \left(\begin{pmatrix} p & h \\ 0 & p \end{pmatrix}, \left(\frac{c}{d}\right) \left(\frac{-pc}{a'}\right) \varepsilon_{a'}^{-1} \varepsilon_d^{-1} \right).$$

Since $dpa' - c'Nb = 1$, we check that $\varepsilon_{a'}\varepsilon_d = \varepsilon_p\left(\frac{-1}{a'}\right)(-1)^{(a'-1)(p-1)/4}$, and that $\left(\frac{c}{a'd}\right) = \left(\frac{c}{p}\right)$. Thus,

$$\left(\frac{c}{d}\right)\left(\frac{-pc}{a'}\right) \Big/ \varepsilon_{a'}\varepsilon_d = \left(\frac{-1}{a'}\right)\left(\frac{p}{a'}\right)\left(\frac{c}{p}\right) \Big/ \varepsilon_p \left(\frac{-1}{a'}\right)(-1)^{(a'-1)(p-1)/4}$$

$$= \varepsilon_p^{-1}\left(\frac{a'c}{p}\right) = \varepsilon_p^{-1}\left(\frac{bhc}{p}\right),$$

since $b \equiv ha' \pmod p$; but $bc \equiv -1 \pmod p$, and so we finally obtain

$$\left(\begin{pmatrix} p & h \\ 0 & p \end{pmatrix}, \varepsilon_p^{-1}\left(\frac{-h}{p}\right) \right).$$

Thus, the contribution of all of the β_h to (3.4) is equal to

$$p^{(k/2)-2}\chi(p) \sum_{h=1}^{p-1} f(z) \left[\left(\begin{pmatrix} p & h \\ 0 & p \end{pmatrix}, \varepsilon_p^{-1}\left(\frac{-h}{p}\right) \right) \right]_{k/2}$$

$$= p^{(k/2)-2}\chi(p)\left(\frac{-1}{p}\right)\varepsilon_p^k \sum_{h=1}^{p-1} \left(\frac{h}{p}\right) f\left(z + \frac{h}{p}\right).$$

Now

$$\sum_{h=1}^{p-1}\left(\frac{h}{p}\right)f\left(z+\frac{h}{p}\right)=\sum_{n=0}^{\infty}a_ne^{2\pi inz}\sum_{h=1}^{p-1}\left(\frac{h}{p}\right)e^{2\pi inh/p}=\varepsilon_p\sqrt{p}\sum_{n=0}^{\infty}\left(\frac{n}{p}\right)a_ne^{2\pi inz},$$

where we have made the usual change of variables (replacing nh by h) and used the value $\varepsilon_p\sqrt{p}$ for the Gauss sum. Thus, we obtain

$$\begin{aligned}p^{(k/2)-2}\chi(p)\left(\frac{-1}{p}\right)\varepsilon_p^{k+1}\sqrt{p}\sum_{n=0}^{\infty}\left(\frac{n}{p}\right)a_ne^{2\pi inz}\\=p^{\lambda-1}\chi(p)\left(\frac{-1}{p}\right)^{\lambda}\sum_{n=0}^{\infty}\left(\frac{n}{p}\right)a_ne^{2\pi inz}.\end{aligned}$$

So the contribution to the n-th coefficient of $T_{p^2}f(z)$ is $p^{\lambda-1}\chi(p)(\frac{(-1)^{\lambda}n}{p})a_n$. This is the middle term on the right in (3.5). The proof of Proposition 13 is complete. □

We note that it can be shown (see Problem 3 below) that the formula in Proposition 13 also holds when $p|N$, in which case $\chi(p)=\chi(p^2)=0$, so we have simply $b_n=a_{p^2n}$.

We further note that, as in the case of integer weight, one can show that the different Hecke operators T_{n^2} commute; that $T_{n^2m^2}=T_{n^2}T_{m^2}$ when g.c.d.$(m, n)=1$; and that $T_{p^{2\nu}}$ is a polynomial in T_{p^2}. Thus, the operators T_{p^2} for different p generate the algebra of operators $\mathbb{C}[\{T_{n^2}\}_{n=1}^{\infty}]$.

In the case of integer weight, we applied a formula analogous to (3.5) in the case when we have a modular form which happens to be an eigenfunction for all of the Hecke operators. The result was a formula for the ratio of a_n to a_1, which can be written in the form $\Sigma a_nn^{-s}=a_1\cdot$(Euler product). (See Propositions 36 and 40 in §III.5.)

In the case of half integer weight, we can consider modular forms which happen to be eigenfunctions for all of the Hecke operators. But since only the T_{n^2} are nontrivial, we only obtain a formula for the ratio of $a_{l_0l_1^2}$ to a_{l_0}, i.e., we can relate coefficients whose indices differ by a perfect square factor.

As in the integer weight case (see the end of §III.5), the spaces $M_{k/2}(\tilde{\Gamma}_0(N), \chi)$ have a basis of eigenforms for all Hecke operators of index prime to N; and certain important subspaces have a basis of eigenforms for *all* of the Hecke operators, i.e., for T_{n^2} when g.c.d.$(n, N)=1$ and for T_p when $p|N$. Thus, let us now suppose that $f\in M_{k/2}(\tilde{\Gamma}_0(N), \chi)$ is an eigenform for all of the Hecke operators T_{n^2}.

Proposition 14. *Let* $f(z)=\Sigma_{n=0}^{\infty}a_ne^{2\pi inz}\in M_{k/2}(\tilde{\Gamma}_0(N), \chi)$ *be an eigenform for all of the Hecke operators* T_{p^2}. *Let* λ_p *be the corresponding eigenvalue, i.e.,* $T_{p^2}f=\lambda_pf$. *Suppose that* l_0 *is not divisible by any square prime to* N *(i.e.,* $p^2|l_0$ *only if* $p|N$*). Then*

$$\sum_{l_1=1}^{\infty} a_{l_0 l_1^2} l_1^{-s} = a_{l_0} \prod_{\text{all } p} \frac{1 - \chi(p)\left(\frac{(-1)^{\lambda} l_0}{p}\right) p^{\lambda-1-s}}{1 - \lambda_p p^{-s} + \chi(p^2) p^{k-2-2s}}. \tag{3.6}$$

(*Note:* The use of the letter λ to denote $(k-1)/2$ and its use with a subscript to denote an eigenvalue should not cause confusion.)

PROOF. If $T_{p^2} f = \lambda_p f$ with $p \nmid N$, then (3.5) gives for any l_1 prime to p:

$$\lambda_p a_{l_0 l_1^2} = a_{l_0 l_1^2 p^2} + p^{\lambda-1}\chi(p)\left(\frac{(-1)^{\lambda} l_0}{p}\right) a_{l_0 l_1^2}; \tag{3.7}$$

$$\lambda_p a_{l_0 l_1^2 p^{2\nu}} = a_{l_0 l_1^2 p^{2(\nu+1)}} + p^{\lambda-1}\chi(p)\left(\frac{(-1)^{\lambda} l_0}{p}\right) a_{l_0 l_1^2 p^{2\nu}} + p^{k-2}\chi(p^2) a_{l_0 l_1^2 p^{2(\nu-1)}}, \tag{3.8}$$

$\nu = 1, 2, \ldots$. If $p | N$, then we need not assume that l_1 is prime to p; we have the same relations (3.7)–(3.8) in all cases, with only the first term on the right nonzero when $p | N$.

On the other hand, if we look at the terms in (3.6) corresponding to all l_1 which differ by a power of p, i.e., if we consider $a_{l_0 l_1^2 p^{2\nu}}(l_1 p^{\nu})^{-s}$ for fixed l_1 prime to p (if $p \nmid N$) and variable $\nu = 0, 1, 2, \ldots$, and if we set $X = p^{-s}$, we find that (3.6) is formally equivalent to the following set of identities for all p and all l_1 prime to p (if $p \nmid N$):

$$\sum_{\nu=0}^{\infty} a_{l_0 l_1^2 p^{2\nu}} X^{\nu} = a_{l_0 l_1^2} \frac{1 - \chi(p)\left(\frac{(-1)^{\lambda} l_0}{p}\right) p^{\lambda-1} X}{1 - \lambda_p X + \chi(p^2) p^{k-2} X^2}. \tag{3.9}$$

But when we multiply both sides of (3.9) by the denominator $1 - \lambda_p X + \chi(p^2)p^{k-2}X^2$ and compare coefficients of $X^{\nu+1}$, we obtain (3.7) for $\nu = 0$ and (3.8) for $\nu = 1, 2, \ldots$. Thus, (3.9) holds, and we have established (3.6). □

Proposition 14 explains the appearance of Euler products of the type we found in the last section (compare (3.6) with (2.31)).

In the next section, we start by formulating Shimura's theorem, which gives a deeper significance to the Euler product in Proposition 14. The Euler product (3.6) turns out to be closely related to a Euler product for a modular form of integral weight $k - 1$.

PROBLEMS

1. Let $\xi = (\alpha, \phi(z)) \in G$, let $\Gamma' \subset \Gamma_0(4)$ be a congruence subgroup, and let $\Gamma'' = \Gamma' \cap \alpha^{-1}\Gamma'\alpha$. For $\gamma \in \Gamma''$ define $t(\gamma)$ by the relation $\xi\tilde{\gamma}\xi^{-1} = \tilde{\gamma}_1 \cdot (1, t(\gamma))$, where $\gamma_1 = \alpha\gamma\alpha^{-1}$.
   (a) Show that the map t depends only on α and not on the $\phi(z)$ in ξ. Prove that t is a group homomorphism from Γ'' to T.
   (b) As usual, let $\tilde{}$ denote the lifting of an element or subgroup of $\Gamma_0(4)$ to G, i.e.,

$\tilde{\gamma} = (\gamma, j(\gamma, z))$ for $\gamma \in \Gamma_0(4)$, $\tilde{\Gamma}' = \{\tilde{\gamma} | \gamma \in \Gamma'\}$ for $\Gamma' \subset \Gamma_0(4)$. Let $K \subset \Gamma''$ be the kernel of t. Show that $\tilde{K} = \tilde{\Gamma}' \cap \xi^{-1}\tilde{\Gamma}'\xi$. Thus, if t is trivial, then $\gamma \mapsto \tilde{\gamma}$ gives an isomorphism from Γ'' to $\tilde{\Gamma}'' = \tilde{\Gamma}' \cap \xi^{-1}\tilde{\Gamma}'\xi$.

(c) Prove that for $f \in M_{k/2}(\tilde{\Gamma}')$, if t is nontrivial, then $f|[\tilde{\Gamma}'\xi\tilde{\Gamma}']_{k/2} = 0$. Recall that $f|[\tilde{\Gamma}'\xi\tilde{\Gamma}']_{k/2} = \Sigma_j f|[\xi\tilde{\gamma}_j]_{k/2}$, where the sum is over all γ_j such that $\tilde{\Gamma}'\xi\tilde{\Gamma}' = \bigcup_j \tilde{\Gamma}'\xi\tilde{\gamma}_j$ (disjoint union).

2. Prove that $t(\gamma) = \left(\frac{n}{d}\right)$ for $\alpha = \left(\begin{smallmatrix} 1 & 0 \\ 0 & n \end{smallmatrix}\right)$, $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(4) \cap \alpha^{-1}\Gamma_0(4)\alpha$.

3. (a) For $\Gamma' = \Gamma_1(N)$, $4|N$, and $\alpha = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 2 \end{smallmatrix}\right)$, show that t is trivial if and only if $8|N$.
(b) For $\Gamma' = \Gamma_1(N)$, $4p|N$, and $\alpha = \left(\begin{smallmatrix} 1 & 0 \\ 0 & p \end{smallmatrix}\right)$, show that t is trivial.
(c) For Γ', p, and α as in part (b), show that $T_{p^\nu} f(z) = \Sigma\, a_{p^\nu n} e^{2\pi i n z}$ for $f(z) = \Sigma\, a_n e^{2\pi i n z} \in M_{k/2}(\tilde{\Gamma}_1(N))$, $\nu = 1, 2, \ldots$. If $p = 2$ and $8 \nmid N$, show that this is still true for ν even. In particular, Proposition 14 holds for $p|N$.

4. Compare the formula for the q-expansion coefficients b_n for $T_{p^2} f$ when $f \in M_{k/2}(\tilde{\Gamma}_0(N), \chi)$ with the corresponding formula for $T_{p^2} f$ when $f \in M_k(N, \chi)$.

5. We noted in the text that T_{p^2} takes $M_{k/2}(\tilde{\Gamma}_1(N))$ to itself. Show that T_{p^2} preserves $M_{k/2}(\tilde{\Gamma}_0(N), \chi)$ for any Dirichlet character χ modulo N.

§4. The theorems of Shimura, Waldspurger, Tunnell, and the congruent number problem

We now state Shimura's fundamental theorem giving a correspondence from forms of half integer weight $k/2$ to forms of (even) integer weight $k - 1$.

Theorem ([Shimura 1973a]). *Let $k \geq 3$ be an odd integer, $\lambda = (k-1)/2$, $4|N$, χ be a Dirichlet character modulo N. Let $f(z) = \Sigma_{n=1}^{\infty} a_n e^{2\pi i n z} \in S_{k/2}(\tilde{\Gamma}_0(N), \chi)$ be an eigenform for T_{p^2} for all primes p with corresponding eigenvalue λ_p: $T_{p^2} f = \lambda_p f$. Define a function $g(z) = \Sigma_{n=1}^{\infty} b_n e^{2\pi i n z}$ by the formal identity*

$$\sum_{n=1}^{\infty} b_n n^{-s} = \prod_{\text{all } p} \frac{1}{1 - \lambda_p p^{-s} + \chi(p)^2 p^{k-2-2s}}. \tag{4.1}$$

Then $g \in M_{k-1}(N', \chi^2)$ for some integer N' which is divisible by the conductor of χ^2. If $k \geq 5$, then g is a cusp form.

Notice that the definition (4.1) of the b_n is equivalent to the following:

(1) $b_1 = 1$;
(2) $b_p = \lambda_p$ for all primes p;
(3) $b_{p^\nu} = \lambda_p b_{p^{\nu-1}} - \chi(p)^2 p^{k-2} b_{p^{\nu-2}}$ for $\nu \geq 2$;
(4) $b_{mn} = b_m b_n$ if m and n are relatively prime.

In Shimura's original theorem, the determination of the level N' of g was a little complicated. However, it has since been shown ([Niwa 1975]) that

one can always take $N' = N/2$. It should also be noted that Shimura actually proved a somewhat more general theorem, applying to f which are not necessarily eigenforms for all of the T_{p^2}.

As a simple numerical example, suppose $N = 4$, χ is trivial. The first nonzero cusp form of half integer weight occurs when $k = 9$, $\lambda = 4$ (see Problem 5 of §IV.1). Up to a constant multiple, it is $f = \Theta F(\Theta^4 - 16F) = \Sigma a_n q^n$, where we have chosen f so that $a_1 = 1$. (By Problem 17(h) of §III.3 this f is also equal to $\eta^{12}(2z)/\Theta^3(z)$.) Clearly, f is an eigenform for all T_{p^2}, because $S_{9/2}(\tilde{\Gamma}_0(4))$ is one-dimensional. Then Shimura's theorem holds with $N' = 2$. Now $S_8(\Gamma_0(2))$ is one-dimensional and spanned by the normalized form $g(z) = (\eta(z)\eta(2z))^8 = q\prod_{n=1}^{\infty}(1 - q^n)^8(1 - q^{2n})^8 = \Sigma b_n q^n$ (see Propositions 19 and 20 in §III.3); hence this $g(z)$ must be the $g(z)$ in the theorem. It is now easy to relate the coefficients b_n of g to the coefficients a_n of f. Namely, using (3.7) with $l_0 = l_1 = 1$, $\lambda = 4$, and noting that $a_1 = 1$, $\chi' = \chi$ (this is the trivial character mod 2, which equals 1 on odd numbers and 0 on even numbers), we obtain:

$$b_p = \lambda_p = a_{p^2} + p^3 \quad \text{if} \quad p > 2; \qquad b_2 = a_4. \tag{4.2}$$

While (4.2) follows immediately from Shimura's theorem, it is nevertheless quite a remarkable numerical identity: the p-th coefficient in $q\prod(1 - q^n)^8 (1 - q^{2n})^8$ is equal to p^3 plus the p^2-th coefficient in $q\prod(1 - q^{2n})^{12}/(\Sigma q^{n^2})^3$! Like many numerical relations that follow from the theory of modular forms, this fact looks rather outlandish when stated in this elementary form without the theoretical context.

If we have a fixed set of linearly independent forms f_i in $S_{k/2}(\tilde{\Gamma}_0(N), \chi)$ which satisfy the hypotheses of Shimura's theorem, then we can extend the Shimura map by linearity to the subspace of $S_{k/2}(\tilde{\Gamma}_0(N), \chi)$ spanned by them. Note that the image g_i of f_i is always a normalized eigenform in $S_{k-1}(N/2, \chi^2)$. If we take another set $\{f_i'\}$ of forms which satisfy Shimura's theorem and are also a basis for the same space as the f_i (for example, if we multiply each f_i by a scalar), the Shimura map is clearly affected. When we refer to the image of a single eigenform under the Shimura map, we shall always mean the normalized eigenform g in Shimura's theorem. But if we have a space of modular forms with fixed basis f_i of eigenforms with Shimura$(f_i) = g_i$, then we define Shimura$(\Sigma a_i f_i) = \Sigma a_i g_i$, which is not necessarily a normalized eigenform. So our meaning of "image" or "preimage" under the Shimura map depends upon the context.

In general, it is possible for several different $f \in S_{k/2}(\tilde{\Gamma}_0(N), \chi)$ to go to the same $g \in M_{k-1}(\Gamma_0(N'), \chi^2)$. For instance, there might be an $f' \in S_{k/2}(\tilde{\Gamma}_0(N), \chi')$ which corresponds to g, where χ' is a character different from χ which has the same square: $\chi^2 = \chi'^2$. However, when $N = 4$, Kohnen [1980] described the situation more precisely. We now describe his main result.

Let $M_{k/2}^{+}(\tilde{\Gamma}_0(4))$ denote the subspace of $M_{k/2}(\tilde{\Gamma}_0(4))$ consisting of $f(z) = \Sigma a_n q^n$ for which $a_n = 0$ whenever $(-1)^\lambda n \equiv 2$ or 3 modulo 4. It is certainly *a priori* possible that there are no nonzero forms f with this property, which

requires that "half" its coefficients vanish. However, we know from Proposition 9 that $H_{k/2}$ is such a form. Also, it is easy to check (using Proposition 17(a) of §III.3 and (1.8) of §IV.1) that $\Theta(z)f(4z) \in M^+_{k/2}(\tilde{\Gamma}_0(4))$ for any $f \in M_{(k-1)/2}(SL_2(\mathbb{Z}))$. It turns out that $M^+_{k/2}(\tilde{\Gamma}_0(4))$ is the direct sum of the one-dimensional space spanned by the Eisenstein series $H_{k/2}$ and the subspace $S^+_{k/2}(\tilde{\Gamma}_0(4))$ of cusp forms:

$$S^+_{k/2}(\tilde{\Gamma}_0(4)) = \{f = \sum a_n q^n \in S_{k/2}(\tilde{\Gamma}_0(4)) | a_n = 0 \text{ if } (-1)^\lambda n \equiv 2 \text{ or } 3 \bmod 4\}. \tag{4.3}$$

It is not hard to show that $M^+_{k/2}(\tilde{\Gamma}_0(4))$ is preserved by all of the Hecke operators T_{p^2} except for T_4. According to Shimura's definition of T_4, one has $T_4 \Sigma a_n q^n = \Sigma a_{4n} q^n$. If we look back at §IV.2, we see that $F_{k/2}$ is an eigenform for T_4 (with eigenvalue 2^{k-2}, see (2.21)), but $H_{k/2}$ is not. In fact, it is easy to see that $T_4 H_{k/2} \notin M^+_{k/2}(\tilde{\Gamma}_0(4))$. For this reason, Kohnen modifies T_4, and defines a slightly different operator T_4^+ so that the m-th coefficient of $T_4^+ \Sigma a_n q^n$ for $(-1)^\lambda m \equiv 2$ or $3 \bmod 4$ is zero and for $(-1)^\lambda m \equiv 0$ or $1 \bmod 4$ is equal to

$$a_{4m} + \chi_{(-1)^\lambda m}(2) 2^{\lambda-1} a_m + 2^{k-2} a_{m/4}.$$

With Shimura's definition, since we have $\chi(p) = 0$ when $p|N$ even if χ is the trivial character on $(\mathbb{Z}/N\mathbb{Z})^*$, the second and third terms vanish in

$$a_{4m} + \chi\chi_{(-1)^\lambda m}(2) 2^{\lambda-1} a_m + \chi(4) 2^{k-2} a_{m/4}.$$

Thus, Kohnen's modification is to replace the trivial character χ by the map which takes the value 1 (and never 0) on all numbers including those not prime to N. It is T_4^+ rather than T_4 which preserves $M^+_{k/2}(\tilde{\Gamma}_0(4))$ and $S^+_{k/2}(\tilde{\Gamma}_0(4))$. Note that $H_{k/2}$ *is* an eigenform for T_4^+ with eigenvalue $1 + 2^{k-2}$.

Kohnen further shows that $M^+_{k/2}(\tilde{\Gamma}_0(4))$ has a basis of eigenforms for all of the T_{p^2} $(p \neq 2)$ and for T_4^+ which is unique up to permutation of the elements and scalar multiplication. There is no obvious way to normalize an eigenform $f = \Sigma a_n q^n$; for example, we cannot necessarily multiply by a scalar to get $a_1 = 1$, since $a_1 = 0$ for all $f \in M^+_{k/2}(\tilde{\Gamma}_0(4))$ if λ is odd. But one can require that the coefficients all lie in as small a field extension of $\mathbb{Q}$ as possible.

It turns out that the images g of these eigen-basis forms $f \in S^+_{k/2}(\tilde{\Gamma}_0(4))$ under the correspondence in Shimura's theorem are all contained in $S_{k-1}(\Gamma)$, $\Gamma = SL_2(\mathbb{Z})$ (the results of Shimura–Niwa only guarantee that they are in $S_{k-1}(\Gamma_0(2))$); they are all distinct; and they form a basis for $S_{k-1}(\Gamma)$ consisting of normalized eigenforms for all of the Hecke operators T_n acting on $S_{k-1}(\Gamma)$. (In particular, $\dim S^+_{k/2}(\tilde{\Gamma}_0(4)) = \dim S_{k-1}(\Gamma)$.) Thus, if we take each of our basis elements for $S^+_{k/2}(\tilde{\Gamma}_0(4))$ to its image under the Shimura map—and take $H_{k/2}$ to the normalized Eisenstein series $-\frac{2(k-1)}{B_{k-1}} E_{k-1}$—and then extend by linearity to all of $M^+_{k/2}(\tilde{\Gamma}_0(4))$, we obtain an isomorphism from $M^+_{k/2}(\tilde{\Gamma}_0(4))$ to $M_{k-1}(\Gamma)$ (and from $S^+_{k/2}(\tilde{\Gamma}_0(4))$ to $S_{k-1}(\Gamma)$). This isomorphism commutes

with the Hecke operators, in the sense that:

$$\begin{array}{ccc} M^+_{k/2}(\tilde\Gamma_0(4)) \ni f & \longmapsto & g \in M_{k-1}(\Gamma) \\ T_{p^2}\downarrow & & \downarrow T_p \\ T_{p^2}f & \mapsto & T_p g \end{array} \qquad \text{and} \qquad \begin{array}{ccc} f & \longmapsto & g \\ T_4^+\downarrow & & \downarrow T_2 \\ T_4^+ f & \mapsto & T_2 g \end{array}$$

The basic method of Shimura's proof of his theorem was to use Weil's theorem which we discussed briefly at the end of §III.3. Weil's theorem says that if $\Sigma b_n n^{-s}$ and its "twists" $\Sigma b_n\psi(n)n^{-s}$ for certain Dirichlet characters ψ each satisfy the right type of functional equation relating the value at s to the value at $k-1-s$, then $g = \Sigma b_n q^n \in M_{k-1}(\Gamma_0(N'), \chi^2)$. But the proof that all of these functional equations are satisfied is not easy; about twenty pages of [Shimura 1973a] are devoted to an investigation of delicate analytic properties of the Dirichlet series corresponding to g and its twists.

Shimura's correspondence seems rather roundabout. It says: take the q-expansion of a suitable $f \in S_{k/2}(\tilde\Gamma_0(N), \chi)$; look at the q-expansion coefficients a_n as n varies over integers with fixed squarefree part l_0, and form a Dirichlet series from them which turns out to have an Euler product; then take the part of this Euler product which is independent of l_0, and expand it into a new Dirichlet series $\Sigma b_n n^{-s}$; and finally, go from this new Dirichlet series to the q-expansion $\Sigma b_n q^n$, which will be your modular form of integral weight.

After Shimura's paper appeared, people started looking for a more conceptual, less roundabout construction of the Shimura correspondence. Certain more direct, analytic constructions were given by Shintani [1975] and Niwa [1975]. In addition, group representation theory was found to provide a conceptual explanation of this correspondence (see [Gelbart 1976], [Flicker 1980]). Moreover, the use of representation theory has led to striking new results about forms of half integer weight, especially in the work of J.-L. Waldspurger.

Using representation theory, Waldspurger [1980, 1981] proved a remarkable theorem establishing a close connection between critical values of L-series for a modular form g of weight $k-1 \in 2\mathbb{Z}$ and the coefficients in the q-expansion of a form f of half integer weight $k/2$ which corresponds to g under the Shimura map. Roughly speaking, the theorem says that the critical value is equal to the square of a corresponding q-expansion coefficient times a nonzero factor which can be explicitly described. Waldspurger's general result is complicated to state, so we shall only describe what it says in two particular situations.

As mentioned before, Kohnen [1980] showed that the Shimura map gives an isomorphism

$$S^+_{k/2}(\tilde\Gamma_0(4)) \xrightarrow{\text{Shimura}} S_{k-1}(\Gamma). \tag{4.4}$$

Here $S^+_{k/2}(\tilde\Gamma_0(4))$ is defined in (4.3). Let $g(z) = \Sigma b_n q^n \in S_{k-1}(\Gamma)$ be a normalized eigenform for all of the Hecke operators, and let χ_D be the character corresponding to the quadratic field of discriminant D. Suppose that

$(-1)^{\lambda}D > 0$, i.e., the quadratic field is real if $\lambda = (k-1)/2$ is even and it is imaginary if λ is odd. Recall that $L_g(\chi_D, s)$ denotes the analytic continuation of $\Sigma_{n=1}^{\infty} \chi_D(n) b_n n^{-s}$ (which can be shown to converge absolutely if $\mathrm{Re}\, s > k/2$). Let $f(z) = \Sigma\, a_n q^n \in S_{k/2}^{+}(\tilde{\Gamma}_0(4))$ be the unique preimage of g under the Shimura map (4.4), i.e., $g = \mathrm{Shimura}(f)$. Let $\langle f, f\rangle$ and $\langle g, g\rangle$ denote the Petersson scalar products, where the same definition (see (5.31) in §III.5) is used for half integer weight as for integer weight, i.e.,

$$\langle f, f\rangle = \frac{1}{6}\int_{F_0(4)} |f(z)|^2 y^{k/2} \frac{dx\,dy}{y^2},$$

where $F_0(4)$ is a fundamental domain for $\Gamma_0(4)$.

Theorem ([Kohnen–Zagier 1981]). *With the above notation and hypotheses,*

$$L_g(\chi_D, \lambda) = \left(\frac{\pi}{|D|}\right)^k \frac{\sqrt{|D|}}{(k-1)!} \frac{\langle g, g\rangle}{\langle f, f\rangle} a_{|D|}^2. \tag{4.5}$$

The basis of eigenforms $f \in S_{k/2}^{+}(\tilde{\Gamma}_0(4))$ can be chosen so that the q-expansion coefficients are all in some totally real number field. However, there is no natural way to normalize them: we can multiply each f by an arbitrary constant c in that field. But note that the right side of (4.5) remains unchanged when f is multiplied by c, since $a_{|D|}^2$ and $\langle f, f\rangle$ are both multiplied by c^2. So (4.5) does not depend on our choice of basis in defining the Shimura isomorphism (4.4).

The L-series value in (4.5) is a "critical value" in the following sense. Recall that the Riemann zeta-function has a functional equation relating $\zeta(s)$ to $\zeta(1-s)$, and the region $0 < \mathrm{Re}\, s < 1$ is called the "critical strip" for $\zeta(s)$. Similarly, the Hasse–Weil L-function of the elliptic curve $E = E_n$ in Chapter II has a functional equation relating $L(E, s)$ to $L(E, 2-s)$, and the region $0 < \mathrm{Re}\, s < 2$ is called its critical strip. The value of such a function at an integer in the critical strip is called a "critical value"; in the case of $L(E, s)$ the critical value is $L(E, 1)$. It is such critical values that have been found to have arithmetic significance. (A general context for the study of critical values is described in [Deligne 1979].) In the case of the L-functions for modular forms g of weight $k-1$, it turns out that they have functional equations relating $L_g(\chi, s)$ to $L_g(\bar{\chi}, k-1-s)$. (Of course, $\chi = \bar{\chi}$ when we are working with quadratic characters $\chi = \chi_D$.) Thus, the critical strip for $L_g(\chi_D, s)$ is $0 < \mathrm{Re}\, s < k-1$; and the critical values are $L_g(\chi_D, j)$ for $j = 1, 2, \ldots, k-2$. The critical value $L_g(\chi_D, \lambda)$ at $j = \lambda = (k-1)/2$ in (4.5) is the value taken at the exact center of the critical strip, i.e., at the fixed point under $s \leftrightarrow k-1-s$.

The first numerical example of the theorem of Kohnen–Zagier occurs when $k = 13$, $\lambda = 6$, since $S_{12}(\Gamma) = \mathbb{C}\Delta$ is the first nonzero space of cusp forms for Γ. In this case $g = \Delta = \Sigma\, \tau(n) q^n$, and $f = \Theta F(\Theta^4 - 16F)(\Theta^4 - 2F)$, where $F = \Sigma_{n\,\mathrm{odd}}\, \sigma_1(n) q^n$, $\Theta = \Sigma\, q^{n^2}$ (see Problem 5(c) in §IV.1; this $f \in S_{13/2}^{+}(\tilde{\Gamma}_0(4))$ can be given other convenient expressions, for example, in

terms of Θ and E_4). For more computational details of this example, see [Kohnen–Zagier 1981].

Our second example of Waldspurger's theorem is the one studied by Tunnell for application to the congruent number problem. In §III.3 we explained that the Hasse–Weil L-function $L(E_1, s) = \Sigma\, b_n n^{-s}$ for the elliptic curve $E_1 : y^2 = x^3 - x$ corresponds to a cusp form $\Sigma\, b_n q^n$ of weight two. In turns out that $g(z) = \Sigma\, b_n q^n$ is in $S_2(\Gamma_0(32))$; g is a normalized eigenform for all of the Hecke operators, and, in fact, is the unique such "new-form", i.e., form which does not come from any lower level $N < 32$.

For n squarefree, let $D = -n$ if $-n \equiv 1 \bmod 4$, $D = -4n$ if $-n \equiv 2$ or $3 \bmod 4$; that is, D is the discriminant of the imaginary quadratic field $\mathbb{Q}(\sqrt{-n})$. In Chapter II we saw that $L(E_n, s) = \Sigma\, \chi_D(m) b_m m^{-s}$ is a twisting of $L(E_1, s)$. (Actually, in (5.7) of §II.5 we wrote $\chi_n(m)$ rather than $\chi_D(m)$; but $b_m = 0$ unless $m \equiv 1 \bmod 4$, in which case $\chi_D(m) = (\frac{-n}{m}) = (\frac{n}{m}) = \chi_n(m)$, so one can equally well use either χ_D or χ_n. Because we will be looking at the critical value at $\lambda = 1$, we want to work with quadratic characters of *imaginary* quadratic fields, i.e., $(-1)^\lambda D = -D > 0$.)

Thus, we can write

$$L(E_1, s) = L_g(s) = \sum b_m m^{-s};$$
$$L(E_n, s) = L_g(\chi_D, s) = \sum \chi_D(m) b_m m^{-s}.$$

We saw that the critical value $L(E_n, 1) = L_g(\chi_D, 1)$ vanishes if and only if n is a congruent number ("only if" here is conditional upon the Birch–Swinnerton-Dyer conjecture). It is this critical value which Waldspurger's theorem provides a means of describing.

Let β denote the "real period" of $E_1 : y^2 = x^3 - x$, which is obtained by integrating dx/y over the segment $[1, \infty)$ where y is real:

$$\beta \underset{\text{def}}{=} \int_1^\infty \frac{dx}{\sqrt{x^3 - x}} = 2.622\cdots.$$

Theorem ([Tunnell 1983]). *There exist a form* $f = \Sigma\, a_m q^m \in S_{3/2}(\tilde{\Gamma}_0(128))$ *and a form* $f' = \Sigma\, a'_m q^m \in S_{3/2}(\tilde{\Gamma}_0(128), \chi_2)$ *such that* Shimura(f) = Shimura$(f') = g = \Sigma\, b_m q^m$ *and*

$$L(E_n, 1) = \begin{cases} \dfrac{\beta}{4\sqrt{n}} a_n^2 & \textit{for } n \textit{ odd;} \\[2ex] \dfrac{\beta}{2\sqrt{n}} a'^2_{n/2} & \textit{for } n \textit{ even.} \end{cases} \tag{4.6}$$

In particular, $L(E_n, 1) = 0$ *if and only if* $a_n = 0$ (*n odd*) *or* $a'_{n/2} = 0$ (*n even*).

Before discussing Tunnell's explicit construction of the forms f and f', we shall discuss how this theorem can be viewed as an analog of the results about $H_{k/2}$ which we proved in §IV.2 (see Proposition 6).

We saw that under the Shimura map $H_{k/2} = \Sigma\, c_n q^n \in M_{k/2}(\tilde{\Gamma}_0(4))$ corresponds to

$$g = -\frac{2(k-1)}{B_{k-1}} E_{k-1} = \frac{1}{2}\zeta(2-k) + \sum \sigma_{k-2}(n) q^n \in M_{k-1}(\Gamma).$$

In Problem 16 of §III.3, we saw that $L_g(s) = \zeta(s)\cdot\zeta(s-(k-2))$ and that $L_g(\chi, s) = L(\chi, s)L(\chi, s-(k-2))$, where the L-functions on the right are Dirichlet L-functions. In particular, setting $\chi = \chi_D$ and $s = \lambda = (k-1)/2$, we obtain

$$L_g(\chi_D, \lambda) = L(\chi_D, \lambda)L(\chi_D, 1-\lambda). \tag{4.7}$$

But, by the functional equation for $L(\chi_D, s)$, we can rewrite $L(\chi_D, \lambda)$ in terms of $L(\chi_D, 1-\lambda)$ as we did in (2.15). We obtain an expression for $L(\chi_D, \lambda)$ as a product of the form $*\cdot L(\chi_D, 1-\lambda)$, where $*$ denotes a nonzero factor involving the gamma-function and powers of π. Substituting in (4.7) and using Proposition 6, we obtain

$$L_g(\chi_D, \lambda) = * \cdot c_{|D|}^2. \tag{4.8}$$

When reformulated in this way, the results of §IV.2 are very similar to the previous two examples of Waldspurger's theorem. As in the Kohnen–Zagier formula, on the left in (4.8) we have the value of the L-function for some $g \in M_{k-1}(\Gamma)$, twisted by χ_D, at the center of its critical strip; on the right we have the square of the corresponding q-expansion coefficient of the form in $M_{k/2}^{+}(\tilde{\Gamma}_0(4))$ which goes to g under the Shimura map. However, the Kohnen–Zagier theorem does not include this case, because g and $f = H_{k/2}$ are not cusp forms (in their formula, one cannot even define $\langle f, f\rangle$ and $\langle g, g\rangle$ except for cusp forms). The case (4.8) is not even included in Waldspurger's general theorem, which also applies only to cusp forms. However, we may think of the results of §IV.2, which were proved in an elementary manner, as a "prototype" for theorems such as those of Waldspurger, Kohnen–Zagier, Tunnell.

Recall that if we could take $\lambda = 1$ in Proposition 6, then the coefficients $c_{|D|} = L(\chi_D, 0)$ of $H_{3/2}$ would be essentially the class numbers of imaginary quadratic fields $\mathbb{Q}(\sqrt{D})$. There is actually an analogy between these critical values and the critical values $L(E_n, 1)$ in (4.6). For elliptic curves, a role analogous to that of the ideal class group of $\mathbb{Q}(\sqrt{D})$ seems to be played by the so-called Tate–Shafarevich group Ш. It is the order of Ш which appears in Birch and Swinnerton-Dyer's conjectural formula for $L(E_n, 1)$ (or for $\lim_{s\to 1}(s-1)^{-r}L(E_n, s)$ if $L(E_n, s)$ has an r-order zero at $s = 1$). Formulas for the "average" value of the order of Ш have been conjectured which are analogous to classical results in analytic number theory for the average value of the class number of imaginary quadratic fields. For more about this, see the papers by Goldfeld *et al.* [1979, 1982].

We now return to Tunnell's theorem, and discuss Tunnell's explicit construction of the modular forms f and f' of weight $\frac{3}{2}$ whose n-th or $(n/2)$-th coefficient gives the value of $L(E_n, 1)$.

Tunnell's first task is to find all $f \in S_{3/2}(\tilde{\Gamma}_0(N), \chi)$ whose image under the Shimura map is the modular form $g \in S_2(\Gamma_0(32))$ corresponding to $L(E_1, s)$. According to [Niwa 1975], any such f has $\mathrm{Shim}(f) \in S_2(\Gamma_0(N/2), \chi^2)$. So one might try taking $N = 64$. However, there is no guarantee that Shimura(f) is not in $S_2(\Gamma_0(N'), \chi^2)$ for N' a proper divisor of $N/2$. For example, we know that any $f \in S^+_{k/2}(\tilde{\Gamma}_0(4))$ has Shimura(f) in $S_{k-1}(\Gamma) = S_{k-1}(\Gamma_0(1))$. So in fact the f we want could be in $S_{3/2}(\tilde{\Gamma}_0(N), \chi)$ for N a multiple of 64. Tunnell computed that, in fact, no f of level 64 maps to g, but that all preimages of g under the Shimura map have level 128.

The character χ must be even, since $S_{k/2}(\tilde{\Gamma}_0(N), \chi) = 0$ for odd χ; it must have conductor dividing 128; and χ^2 must be trivial, i.e., χ must be quadratic. There are two such χ: the trivial character $\chi = 1$ and the character χ_2 defined by $\chi_2(j) = (\frac{2}{j})$ for j odd. Tunnell determined that Shimura$^{-1}(g)$ consists of two forms in $S_{3/2}(\tilde{\Gamma}_0(128))$ and two forms in $S_{3/2}(\tilde{\Gamma}_0(128), \chi_2)$. Moreover, he found that these four forms of weight $\frac{3}{2}$ can be constructed in a particularly simple way: by multiplying a certain form f_1 of weight 1 by forms of weight $\frac{1}{2}$ of the type $\Theta(mz)$.

Up to a constant multiple, $\Theta(mz)$ is equal to $\Theta|[(\begin{smallmatrix} m & 0 \\ 0 & 1 \end{smallmatrix}), m^{1/4}]_{1/2}$, and so it easily follows from Problem 1 in §IV.1 that $\Theta(mz) \in M_{1/2}(\tilde{\Gamma}_0(4m), \chi_m)$ (see also Proposition 17 in §III.3). If we multiply this by a form $f_1(z) \in M_1(\Gamma_0(128), \chi)$, then the product $f_1(z)\Theta(mz)$ is contained in $M_{3/2}(\tilde{\Gamma}_0(128), \chi\chi_{-m})$ when $4m|128$. (Recall that $M_1(\Gamma_0(128), \chi) = M_{2/2}(\tilde{\Gamma}_0(128), \chi \cdot \chi_{-1})$ by Proposition 3 of §IV.1, and so the character for $f_1(z)\Theta(mz)$ is $\chi \cdot \chi_{-1} \cdot \chi_m$.)

The form f_1 is chosen to be

$$f_1(z) = \sum_{m,n \in \mathbb{Z}} (-1)^n q^{(4m+1)^2 + 8n^2}.$$

It is an easy exercise to show that

$$f_1(z) = (\Theta(z) - \Theta(4z))(\Theta(32z) - \tfrac{1}{2}\Theta(8z)),$$

and therefore $f_1 \in M_{1/2}(\tilde{\Gamma}_0(16)) \cdot M_{1/2}(\tilde{\Gamma}_0(128), \chi_2) \subset M_1(\Gamma_0(128), \chi_{-2})$. Actually, f_1 also vanishes at the cusps of $\Gamma_0(128)$, and so we have $f_1 \in S_1(\Gamma_0(128), \chi_{-2})$. Thus, for $m|32$ we have

$$f_1(z)\Theta(mz) \in S_{3/2}(\tilde{\Gamma}_0(128), \chi_{2m}).$$

It can be shown, by the way, that f_1 can also be written as a product

$$f_1(z) = \eta(8z)\eta(16z) = q \prod (1 - q^{8n})(1 - q^{16n}).$$

For a short proof using the Jacobi triple product formula, see [Moreno 1980].

Proposition ([Tunnell 1983]). *The modular forms $f_1(z)\Theta(2z)$, $f_1(z)\Theta(8z) \in S_{3/2}(\tilde{\Gamma}_0(128))$ and $f_1(z)\Theta(4z)$, $f_1(z)\Theta(16z) \in S_{3/2}(\tilde{\Gamma}_0(128), \chi_2)$ are a maximal set of linearly independent eigenforms for all of the T_{p^2} whose image under the Shimura map is the modular form $g = \Sigma b_n q^n \in S_2(\Gamma_0(32))$ corresponding to $L(E_1, s)$.*

Tunnell then proves his theorem by a close examination of what Waldspurger says in the very special circumstances of the congruent number problem. In our situation, Waldspurger's theorem boils down to the assertions that: (1) there is a linear combination $f = \Sigma a_n q^n$ of the preimages in $S_{3/2}(\tilde{\Gamma}_0(128))$ such that for all odd squarefree l_0

$$L_g(\chi_{-l_0}, 1) = ca_{l_0}^2$$

for some constant c; (2) there is a linear combination $f' = \Sigma a'_n q^n$ of the preimages in $S_{3/2}(\tilde{\Gamma}_0(128), \chi_2)$ such that for all odd squarefree l_0

$$L_g(\chi_{-2l_0}, 1) = c' a'^2_{l_0}$$

for some constant c'. Tunnell computes that one can take $f(z) = f_1(z)\Theta(2z)$, $c = \beta/4\sqrt{l_0}$, and $f'(z) = f_1(z)\Theta(4z)$, $c' = \beta/2\sqrt{2l_0}$. Since $L_g(\chi_{-n}, s) = L(E_n, s)$, this gives his theorem with

$$f(z) = (\Theta(z) - \Theta(4z))(\Theta(32z) - \tfrac{1}{2}\Theta(8z))\Theta(2z),$$
$$f'(z) = (\Theta(z) - \Theta(4z))(\Theta(32z) - \tfrac{1}{2}\Theta(8z))\Theta(4z).$$

The numerical identity (4.6) in Tunnell's theorem is quite bazaar. By the formula (6.8) in §II.6, it says that for n odd and squarefree:

$$8\sqrt{n}\sum_{m=1}^{\infty}\left(\frac{n}{m}\right) b_m e^{-\pi m/n\sqrt{8}} = a_n^2 \int_1^{\infty} \frac{dx}{\sqrt{x^3 - x}},$$

where $L(E_1, s) = \Sigma b_m m^{-s}$ and a_n is the n-th q-expansion coefficient in (4.9) below!

Note that in (4.6) we are interested only in the odd q-expansion coefficients of f and f'. But for n odd, the n-th coefficient is the same as the n-th coefficient in

$$\Theta(z)\left(\Theta(32z) - \frac{1}{2}\Theta(8z)\right)\Theta(2z) = \sum_{x,y,z\in\mathbb{Z}} q^{2x^2+y^2+32z^2} - \frac{1}{2}\sum_{x,y,z\in\mathbb{Z}} q^{2x^2+y^2+8z^2} \tag{4.9}$$

and

$$\Theta(z)\left(\Theta(32z) - \frac{1}{2}\Theta(8z)\right)\Theta(4z) = \sum_{x,y,z\in\mathbb{Z}} q^{4x^2+y^2+32z^2} - \frac{1}{2}\sum_{x,y,z\in\mathbb{Z}} q^{4x^2+y^2+8z^2}, \tag{4.10}$$

respectively.

(Notice that the coefficient of q^{l_0} in (4.9) and (4.10) is obviously zero if $l_0 \equiv 5$ or $7 \bmod 8$ for (4.9), $l_0 \equiv 3 \bmod 4$ for (4.10); but this tells us nothing new, since in Proposition 12 of §II.6 we saw that $L(E_n, 1) = 0$ if $n = l_0 \equiv 5$ or $7 \bmod 8$ or $n = 2l_0 \equiv 6 \bmod 8$.)

Collecting coefficients of q^{l_0} in (4.9) and (4.10), we conclude the version of Tunnell's theorem that we cited at the beginning of Chapter I.

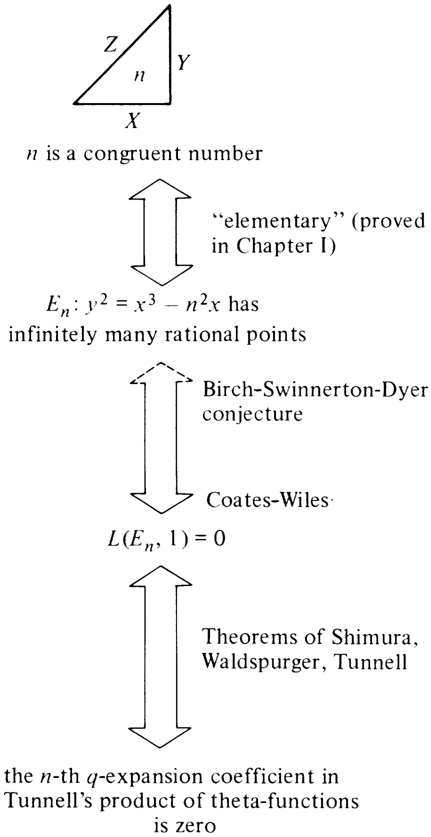


Figure IV.1

Theorem ([Tunnell 1983]). *If n is a squarefree and odd (respectively, even) positive integer and n is the area of a right triangle with rational sides, then*

$$\#\{x, y, z \in \mathbb{Z} \mid n = 2x^2 + y^2 + 32z^2\} = \tfrac{1}{2}\#\{x, y, z \in \mathbb{Z} \mid n = 2x^2 + y^2 + 8z^2\}$$

(*respectively,*

$$\#\left\{x, y, z \in \mathbb{Z} \,\middle|\, \frac{n}{2} = 4x^2 + y^2 + 32z^2\right\} = \frac{1}{2}\#\left\{x, y, z \in \mathbb{Z} \,\middle|\, \frac{n}{2} = 4x^2 + y^2 + 8z^2\right\}\Bigg).$$

If the weak Birch–Swinnerton-Dyer conjecture is true for the elliptic curves E_n: $y^2 = x^3 - n^2x$, then, conversely, these equalities imply that n is a congruent number.

In Fig. IV.1 we recall the logical structure of the argument.

As mentioned in Chapter I, Tunnell's theorem has the practical value of leading to an effective and rapid algorithm for determining whether n is a congruent number. In addition, one can give quick new proofs of certain conditions for n not to be a congruent number. For example, if n is a prime congruent to 3 modulo 8, Tunnell shows that $a_n \equiv 2 \bmod 4$, and therefore n is not a congruent number. For this and other corollaries, see Tunnell's paper.

The only sense in which Tunnell's theorem is not yet a completely satisfactory solution to the ancient congruent number problem is that in one direction it is conditional upon the weak Birch–Swinnerton-Dyer conjecture for certain elliptic curves. But lately, significant progress has been made toward a proof of that conjecture in enough generality to include the curves E_n. As mentioned in §II.6, in the mid-1980s Gross and Zagier were able to show that the weak Birch-Swinnerton-Dyer conjecture is true for E_n for a large class of n (but one must be able to show that $L'(E_n, 1) \neq 0$). At about the same time, R. Greenberg showed that, if the conjecture were to fail for an elliptic curve such as E_n which has complex multiplication, then that would imply a highly improbable combination of consequences for the Tate–Shafarevich group of the elliptic curve (see [Greenberg 1983]).

The next few years saw a series of striking developments which, while not bearing directly on the congruent number problem, brought us closer to having a proof of the Birch–Swinnerton-Dyer conjecture. First, K. Rubin proved that if the group of rational points of an elliptic curve E with complex multiplication has rank at least 2, then its L-function $L(E, s)$ has a zero at $s = 1$ of order at least 2 (see [Rubin 1987]). In combination, the results of Coates–Wiles, Gross–Zagier, and Rubin imply the following: *for an elliptic curve with complex multiplication, if the order of zero $L(E, s)$ at $s = 1$ is either* 0 *or* 1, *then the order of zero is in fact equal to the rank of the group of rational points.* Then Kolyvagin was able to strengthen this result dramatically, by proving that the same theorem is true for the much broader class of modular elliptic curves. If the Taniyama–Weil conjecture is true, then this class includes all elliptic curves defined over the rational numbers. See [Kolyvagin 1989 and 1990] and [Rubin 1989]. (For a readable overview of results on the Birch–Swinnerton-Dyer conjecture, together with a discussion of the relation of this work to Gauss' class number conjecture, see [Ireland and Rosen 1990, Ch. 20].)

It is remarkable that the nearly complete solution that we now have to such an old and naive question as the congruent number problem, has required some of the most powerful and sophisticated tools from diverse branches of twentieth century mathematics.

Answers, Hints, and References for Selected Exercises

§I.1

1. See [Hardy and Wright 1960, pp. 190–191]. **3.** (b) Follow the proof that $x^4 + y^4 = u^2$ is unsolvable on pp. 191–192 of Hardy and Wright. **4.** For fixed n and fixed x (so that Z is fixed), the triples that correspond to x come from the intersection of the two conic sections $X^2 + Y^2 = Z^2$ and $XY = 2n$ in the XY-plane. Given one point of intersection (X, Y), the other three are $(-X, -Y)$, (Y, X), and $(-Y, -X)$, and so do not give a distinct triple. **5.** (a) 1681/144; (b) 25/4; (c) 841/4, 1369/4. **7.** Since $x^2, y^2 \equiv 0$, 1 or 4 (mod 8), it follows that $2x^2 + y^2 + 8z^2$ can *never* equal an integer $n \equiv 5$ or 7 (mod 8). The first congruent number $n \equiv 1$ or 3 (mod 8) is 41, which is the area of the right triangle with sides $6\frac{3}{20}$, $13\frac{1}{3}$, $14\frac{41}{60}$.

§I.2

1. replace y by y/n^2 and x by x/n **2.** (c) $x = -nY/(X + Z)$, $y = 2n^2/(X + Z)$

(e)

X	Y	Z	x	y	X	Y	Z	x	y
3	4	5	−3	9	3	4	−5	12	−36
4	3	5	−2	8	4	3	−5	18	−72
−3	−4	5	12	36	−3	−4	−5	−3	−9
−4	−3	5	18	72	−4	−3	−5	−2	−8

3. (a) If Z is the side opposite θ and X, Y are the other two sides, the law of cosines gives $X^2 + Y^2 - 2AXY = Z^2$. Then the point $u = (X - AY)/Z$, $v = Y/Z$ is on the ellipse $u^2 + B^2v^2 = 1$. Again use the slope of the line joining $(-1, 0)$ to (u, v) to parametrize this ellipse. Show that $u = (1 - B^2t^2)/(1 + B^2t^2)$, $v = 2t/(1 + B^2t^2)$. Now the area of the triangle is $\frac{1}{2}XY \sin\theta = \frac{B}{2}XY$, so we obtain:
$n = \frac{1}{2}BZ^2(Y/Z)(X/Z) = \frac{1}{2}BZ^2v(u + Av) = BZ^2t(1 + 2At - B^2t^2)/(1 + B^2t^2)^2$.
Finally, set $x = -Bt$, $y = (1 + B^2t^2)/Z$, so that $ny^2 = x(x^2 + 2\frac{A}{B}x - 1)$. (b) Since

$A = (1 - \lambda^2)/(1 + \lambda^2)$ and $B = 2\lambda/(1 + \lambda^2)$, the right side of the cubic equation in part (a) becomes $x\left(x^2 + \dfrac{1 - \lambda^2}{\lambda}x - 1\right) = x(x - \lambda)(x + \frac{1}{\lambda})$.

§I.3

1. Counterexample if $K = \mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$: $F(x, y, z) = x^p - x$. **3.** 1; 0; 2 **5.** (b) Here's a counterexample if char $K = p$. Let d be a multiple of p, $F(x, y, z) = x^d + y^d + z^d + x^a y^b z^c$, where $0 < a, b, c,\ a + b + c = d$. Then all partial derivatives vanish at (0, 0, 1), (0, 1, 0), (1, 0, 0), but these points are not on the curve. In fact, if $p = d = 3$, $a = b = c = 1$, then the curve is smooth at all of its points, even though there are three points of $\mathbb{P}_K^2$ where all three partial derivatives vanish. (d) With $K = \mathbb{R}$ or $\mathbb{C}$, think of F as a map $F: K^3 \to K$, so that the smoothness condition becomes nonvanishing of the gradient. Then apply the chain rule to the composite function: $x'y'z'$-space $\xrightarrow{A^{-1}}$ xyz-space $\xrightarrow{F} K$. **6.** (b) Reduce to the case $z_1 = z_2 = 1$, set $\Delta x = \frac{t}{1+t}(x_2 - x_1)$, so that $f'(x_1)\Delta x = \frac{t}{1+t}(y_2 - y_1)$. Then $0 = \tilde{F}(x_1 + \Delta x, y_1 + f'(x_1)\Delta x + a_m \Delta x^m + \cdots, 1) = \tilde{F}(x_1 + \Delta x, y_1 + f'(x_1)\Delta x, 1) + \frac{\partial \tilde{F}}{\partial y}(x_1 + \Delta x, y_1 + f'(x_1)\Delta x, 1) a_m \Delta x^m + \cdots = (1 + t)^{-\deg \tilde{F}} \tilde{F}(x_1 + x_2 t, y_1 + y_2 t, z_1 + z_2 t) + (\text{nonzero constant}) t^m +$ higher terms.

§I.5

2. (a) $f(z) = (e^{2\pi i a}, e^{2\pi i b})$ for $z = a\omega_1 + b\omega_2$, $a, b \in \mathbb{R}$. (b) N^2. (c) In part (a), let $a = j/p$, $b = k/p$; then for $j, k \in \mathbb{F}_p$ not both zero we have $(j, k) \leftrightarrow$ subgp gen by $(e^{2\pi i j/p}, e^{2\pi i k/p})$ gives the required one-to-one correspondence; there are $p + 1$ subgroups. **3.** (a) If $s = 2$, reduce to the case when $f(m, n) = 1$ if $m, n \equiv m_1, n_1 \pmod N$, $f(m, n) = -1$ if $m, n \equiv m_2, n_2 \pmod N$, $f(m, n) = 0$ otherwise (where m_i, n_i are fixed pairs); then pair together the $jN + m_1$, $kN + n_1$ term and the $jN + m_2$, $kN + n_2$ term. (b) $\dfrac{(-N)^{-s}}{(s-1)!} \sum_{0 \le m, n < N} f(m, n) \wp^{(s-2)}(\frac{m}{N}\omega_1 + \frac{n}{N}\omega_2; \omega_1, \omega_2)$.

§I.6

2. $g_2^3 - 27g_3^2$. **3.** $\wp''(z) = 6\wp(z)^2 - \frac{1}{2}g_2$. **6.** $\zeta(8) = \pi^8/9450$. **7.** 4/3, 8/27
10. Set $v(z) = f^{-1}(\wp(z))$, find $(dv/dz)^2$, and show that $dv/dz = \pm 1$.
14. (a) Substitute $t = \sin^2 \theta$ and integrate by parts. (b) Use Problem 12(b), and substitute $x = (t - e_1)/(e_2 - e_1)$ to get the expression under the radical in the form $x(x - 1)(x - \lambda)$. (Note: An elliptic curve written in the form $y^2 = x(x - 1)(x - \lambda)$ is said to be in "Legendre form".) Then make the substitution $t = 1/x$. (c) Expand $(1 - \lambda t)^{-1/2}$ in a binomial series in part (b), and use part (a).

§I.7

2. (a) $((x^2 + n^2)/2y)^2$. (c) let ord_2 denote the power of 2 dividing the numerator minus the power dividing the denominator; dividing into the cases $\text{ord}_2 x < \text{ord}_2 n$, $\text{ord}_2 x = \text{ord}_2 n$, $\text{ord}_2 x > \text{ord}_2 n$, determine $\text{ord}_2(x^2 + n^2)$ and use $y^2 = (x^2 - n^2)x$ to find $\text{ord}_2 y$, in order to conclude that $\text{ord}_2 y \ge \text{ord}_2(x^2 + n^2)$. (Of course, $\text{ord}_2 n = 0$

or 1, since n is squarefree.) **3.** (a) 0 together with 3 x-intercepts. (b) points of inflection. (c) from each of the 3 x-intercepts there are 4 lines which are tangent to the curve at points of order 4. (d) if we have a configuration of three lines crossing three other lines, and if the elliptic curve passes through 8 of the 9 points where they cross, then it passes through the ninth. **4.** (a) eight; setting $y'' = 0$ after twice differentiating $y^2 = f(x)$ implicitly and then multiplying both sides by $2y^2$ gives: $(2yy')^2 = 2y^2 f''(x)$, and hence $f'(x)^2 - 2f(x)f''(x) = 0$. (b) $x = \pm n\sqrt{1 \pm 2\sqrt{3}/3}$. **5.** four; draw lines from $-Q$ which are tangent to the curve. **7.** 4; 3; 8. The four points of order 4 not of order 2 are found by drawing lines from $(n, 0)$ which are tangent to the curve. **8.** two points of order 2 (at infinity and $(a^{1/3}, 0)$); three points of order 3 (at infinity and $((4a)^{1/3}, \pm(3a)^{1/2})$ if a is positive, $(0, \pm(-a)^{1/2})$ if a is negative); four points of order 4, namely, the two of order 2 and the two points $(a^{1/3}(1 + \sqrt{3}), \pm(3a(3 + 2\sqrt{3}))^{1/2})$ (if a is positive; change the two +'s to − if a is negative).

§I.8

1. $\wp(2z) = \dfrac{16x^4 + 8g_2x^2 + 32g_3x + g_2^2}{16(4x^3 - g_2x - g_3)}$, with $x = \wp(z)$.
2. $f_3(z) = 3x^4 - \frac{3}{2}g_2x^2 - 3g_3x - \frac{1}{16}g_2^2$, with $x = \wp(z)$. **3.** Look at zeros and poles of $\wp(Nz) - \wp(z)$; determine the constant -1 by comparing coefficients of z^{-2}. (See [Lang 1978b, pp. 34–35].) **4.** Considered on the points $\frac{j}{N}\omega_1 + \frac{k}{N}\omega_2$ in the z-plane, σ must take every such point either to itself or its negative (modulo L). When finding the matrix entries by looking at the cases $(j, k) = (1, 0)$ or $(0, 1)$, we first obtain $\left(\begin{smallmatrix}\pm 1 & 0\\ 0 & \pm 1\end{smallmatrix}\right)$, but consideration of other j, k shows that both signs must be the same. The analogous situation in cyclotomic fields is to set $\mathbb{Q}_N^+ = \mathbb{Q}(\cos\frac{2\pi}{N})$ (i.e., adjoin just the x-coordinate of the point of order N). Then $\mathrm{Gal}(\mathbb{Q}_N/\mathbb{Q}_N^+)$ is the subgroup $\{\pm 1\}$ in $(\mathbb{Z}/N\mathbb{Z})^*$. **5.** The image is conjugated by the change of basis matrix. **6.** (a) a subgroup of order 2, (b) the trivial subgroup, (c) the entire group, (d) a subgroup of order 2. (Note: In this and the next problem, these subgroups are only defined up to conjugation; see Problem 5.) **7.** (a) 48, (b) Using Problem 4(b) of §I.7, we see that K_3 is generated by $\pm\sqrt{1 \pm 2\sqrt{3}/3}$ and by the solutions y of $y^2 = xn^2(2\sqrt{3}/3) = \pm\frac{2n^3}{3}\sqrt{3 \pm 2\sqrt{3}}$. (c) By part (a), $[K_3 : \mathbb{Q}]$ divides 48. Since it is obtained by successive extraction of square roots, K_3 has degree in fact dividing 16. On the other hand, by part (b), it is easy to see that K_3 contains i and also the fourth root of $4n^2(3 + 2\sqrt{3})$, which satisfies $x^8 - 24n^2x^4 - 48n^4$, which is irreducible (by Eisenstein's criterion for the prime 3, if $3\nmid n$; if $3|n$, a generalization of Eisenstein's criterion can be used). Thus, the field F obtained by adjoining this root has degree 8. But $F \subset \mathbb{R}$, while $K_3 \ni i$, so that $[K_3 : \mathbb{Q}] \geq 16$. Hence, the degree is 16, and the image of the galois group is a 2-Sylow subgroup of $GL_2(\mathbb{Z}/3\mathbb{Z})$. Here are two alternate ways of showing that $[K_3 : \mathbb{Q}]$ is at least 16: (i) K_3 has at least degree 2 over $\mathbb{Q}(\sqrt{2n\sqrt{3 + 2\sqrt{3}}}) \subset \mathbb{R}$, so it suffices to show irreducibility over $\mathbb{Q}$ of the polynomial $\prod_{8\text{ choices of }\pm}(x \pm \sqrt{\pm 2n\sqrt{3 \pm 2\sqrt{3}}})$. But otherwise a product of 4 of the roots would be in $\mathbb{Z}$, and this can be ruled out directly. (ii) First show that $[\mathbb{Q}(\sqrt{3 + 2\sqrt{3}}) : \mathbb{Q}] = 4$, after which it suffices to show that $\mathbb{Q}(\sqrt{2n\sqrt{3 + 2\sqrt{3}}}) \neq \mathbb{Q}(\sqrt{3 + 2\sqrt{3}})$. Otherwise we would obtain $2n\sqrt{3 + 2\sqrt{3}} = (a + b\sqrt{3 + 2\sqrt{3}})^2$,

$a, b \in \mathbb{Q}(\sqrt{3})$; since $\sqrt{3 + 2\sqrt{3}} \notin \mathbb{Q}(\sqrt{3})$, this gives $0 = a^2 + b^2(3 + 2\sqrt{3})$, a contradiction, since $a, b \in \mathbb{R}$. (d) $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)$, i.e., $z_1 = \omega_1/3$, $z_2 = (\omega_1 + \omega_2)/3$. **8.** (c) Show that α is an eigenvalue of a 2×2 matrix with integer entries. **9.** (a) 3, (b) 4, (c) 6, (d) 7, (e) 8, (f) 5, (g) 7.

§I.9

1. Possible conditions on p and f: $8|f$; or else $2|f$ and $p \equiv -1 \pmod{12}$.

2. (a)

p	3	5	7	11	13	17	19	23
type	(2, 2)	(2, 4)	(2, 4)	(2, 2, 3)	(2, 4)	(4, 4)	(2, 2, 5)	(2, 4, 3)

(b) (4, 4) (c) (2, 2, 7), (d) (2, 4, 9), (e) (2, 2, 9, 37). For $q = 17$ and 9, you have to check that the x-coordinates of the doubles of all 12 points not of order 2 are 0, 1 or -1, i.e., all points are of order 4. Note that those cases also follow from Problem 11 below. **3.** Same as Problem 2 except for: $p = 13$, (2, 2, 5); $p = 17$, (2, 2, 5). **4.** Notice that the right side $x^3 - a$ runs through $\mathbb{F}_q$ as x runs through $\mathbb{F}_q$, so the number of points is the same as on $y^2 = x$ (plus the point at infinity). **5.** See Problem 10 below.

6.

p	5	7	11	13	17	19	23
type	(2, 3)	(2, 2)	(4, 3)	(2, 2, 3)	(2, 9)	(2, 2, 7)	(8, 3)

7. Show that for all but finitely many primes $p \equiv -1 \pmod 6$ the group homomorphism from the subgroup of points of order m in $E(\mathbb{Q})$ to $E(\mathbb{F}_p)$ is an imbedding. Show that this implies that points of finite order can only come from 2 points of order 2 and/or 3 points of order 3. Then find whether the points of order 2 or 3 can have rational coordinates. **9.** (a) Show that a point of order N is taken to another point of order N by the complex multiplication automorphism; but if both (x, y) and $(-x, \sqrt{-1}y)$ have coordinates in $\mathbb{F}_q$, then $\sqrt{-1} \in \mathbb{F}_q$. **10.** Proceed as in 9(a), using the complex multiplication $(x, y) \mapsto (\zeta x, y)$, where ζ is a nontrivial cube root of unity in $\mathbb{F}_{q^2}$. **11.** Suppose that $\alpha < \beta$. Let G be the quotient group of the group of points of order l^β modulo the subgroup of points of order l^α. Then $G \approx \mathbb{Z}/l^{\beta-\alpha}\mathbb{Z}$. Show that the complex multiplication used in 9(a) gives an automorphism of the group G whose square is the automorphism -1 (which takes every element in $\mathbb{Z}/l^{\beta-\alpha}\mathbb{Z}$ to its negative). Show that there is no such automorphism if $l \equiv 3 \pmod 4$, because -1 is not a square in $(\mathbb{Z}/l\mathbb{Z})^*$. If $l = 2$, use the fact that -1 is not a square in $(\mathbb{Z}/4\mathbb{Z})^*$.

§II.1

4. $Z(T) = (1 - T^2)^{-1/2}$; $Z(T) = (1 - T^2)^{-1}$; take the equation $x^2 - a = 0$ for a any quadratic nonresidue mod p. **5.** $Z(T) = 1/((1 - T)(1 - 2T) \cdots (1 - (M - 1)T))$. **6.** (a) $Z(T) = 1/[(1 - T)^l(1 - T^l)^{l-1}(1 - T^{l^2})^{l-1} \cdots (1 - T^{l^{M-1}})^{l-1}]$ (b) The limit of the $Z(T)$ in part (a) as M approaches infinity, an infinite product which is not a rational function.
7. (b)–(c) $Z(\mathbb{P}^m_{\mathbb{F}_q}/\mathbb{F}_q; T) = 1/(1 - T)(1 - qT)(1 - q^2T) \cdots (1 - q^mT)$. **9.** Write V as a disjoint union of affine varieties, and use Problem 1. To reduce to the case of a single equation, use induction and the observation that the number of simultaneous

zeros of $f = g = 0$ is equal to the sum of the number for $f = 0$ and the number for $g = 0$ minus the number for $fg = 0$. **10.** $1/(1 - T)(1 - qT)$.
11. $(1 - qT)/(1 - q^2T)(1 - q^3T)$.
12. $1/(1 - T)(1 - qT)(1 - q^2T)^2(1 - q^3T)(1 - q^4T)$. **13.** Following Problem 13 of §I.9, you quickly see that the field extension generated by the points of order l^{M+1} is contained in $\mathbb{F}_{q^{l^M}}$. It remains to show that only the l^M-order points, and none of exact order l^{M+1}, have coordinates in $\mathbb{F}_{q^{l^{M-1}}}$. Prove this by computing the exact power of l that divides $N_{l^{M-1}}$. Consider separately the cases (i) l remains prime in the quadratic extension $\mathbb{Q}(\alpha)$; (ii) the ideal $(l) = L\bar{L}$ splits into a product of two prime ideals. Show that, for example, if L^e is the highest power of L dividing $\alpha - 1$, then L^{e+j} is the highest power of L dividing $\alpha^{l^j} - 1$. In this way, use the fact that N_1 is exactly divisible by l^2 to prove that $N_{l^{M-1}}$ is exactly divisible by l^{2M}. **16.** Write $Z(T) = 1 + c_1T + c_2T^2 + \cdots = P(T)/Q(T)$, $Q(T) = b_0 + b_1T + \cdots$, where $P(T)$ and $Q(T)$ have no common factors and all coefficients are integers. Check that the polynomial $Q(T)$ is primitive. Use the Euclidean algorithm to write $PU + QV = m$, where $U, V \in \mathbb{Z}[T]$, $m \in \mathbb{Z}$, $m \neq 0$. Write $m/Q = U(P/Q) + V = d_0 + d_1T + \cdots$. Since $m = Q(T)(d_0 + d_1T + \cdots)$ with $Q(T)$ primitive, it follows by the proof of Gauss's lemma that $m|d_j$. In particular, $m|d_0$, and this means that the constant term b_0 is ± 1. It immediately follows that $P(T)$ also has constant term ± 1.

§II.2

1. For example, to prove (3), in the double sum for $J(\chi_1, \chi_2)g(\chi_1\chi_2)$, replace x by x/y and then replace y by $x + y$. **4.** $i\sqrt{3}, \sqrt{5}, i\sqrt{7}, 3$. **5.** $1 \pm 2i, 3, -3 \pm 2i, -1 \pm 4i$.
6. $\chi(4)J(\chi, \chi) = \Sigma\,\chi(4x - 4x^2) = \Sigma\,\chi(1 - (2x - 1)^2) = \Sigma\,\chi(1 - x^2)$ (replacing $2x - 1$ by x) $= \Sigma\,(1 + \chi_2(y))\chi(1 - y)$ (where $y = x^2$ if $\chi_2(y) = 1$), and this equals $J(\chi_2, \chi)$. **7.** $(1 \pm 3i\sqrt{3})/2, (-5 \pm 3i\sqrt{3})/2$. **8.** (a) Use part (b) with m a square root of $1/n$ in $\mathbb{F}_{q^2}$. (b) Replace x by x/m^2 and y by y/m^3 in $y^2 = x^3 - n^2x$. **9.** (a) Replace x by x/a. (b) Choose J to be the ideal of elements x for which $ax \in I$, and take the sum over a fixed coset in $(R/I)^*$ of the subgroup consisting of elements congruent to 1 modulo J; show that each such sum vanishes. (Look at the example $R = \mathbb{Z}$, $I = (N)$ to get used to the argument.) (c) Check that $\overline{g(\chi, \psi)} = g(\bar{\chi}, \bar{\psi}) = \chi(-1)g(\bar{\chi}, \psi)$. Then we have $|g(\chi, \psi)|^2 = g(\chi, \psi)g(\bar{\chi}, \bar{\psi}) = \Sigma_{x \in (R/I)^*}\Sigma_{y \in R/I}\,\chi(x)\bar{\chi}(y)\psi(x - y)$ (here it makes no difference whether we sum over R/I or $(R/I)^*$). Replace y by xy in the inner sum, thereby changing the summand to $\bar{\chi}(y)\psi(x(1 - y))$. For fixed x *not* prime to I, if we let J denote the ideal of elements whose product with x is in I (thus, J is strictly larger than I), we see that the inner sum vanishes by the argument in part (b). So we can replace the outer sum $\Sigma_{x \in (R/I)^*}$ by $\Sigma_{x \in R/I}$. We then interchange the order of summation to obtain: $|g(\chi, \psi)|^2 = \Sigma_{y \in R/I}\,\bar{\chi}(y)\,\Sigma_{x \in R/I}\,\psi(x(1 - y))$. When $y = 1$, the inner sum is $\mathbb{N}I$. But when $y \neq 1$ it vanishes, since, by assumption, ψ is nontrivial on the subgroup $(1 - y)R + I/I$ in R/I. **10–17.** See Weil's paper [Weil 1949] or else [Ireland and Rosen 1982, §4 in Chapter 11]. **20.** (b) By part (a), modulo 3 we have $J(\chi_3, \chi_3) \equiv (\Sigma\chi_3(x)\psi(x))^3 \equiv \Sigma\chi_3^3(x)\psi^3(x) = \Sigma_{x \in \mathbb{F}_q^*}\psi(3x) = -1$. (c) $J(\chi_3, \chi_3)$ is an element of $\mathbb{Z}[\omega]$ of norm p which is congruent to -1 mod 3. (e) Since $J(\chi_3, \chi_3)$ has norm p and the right congruence, it suffices to show that it is in the ideal $(a + b\omega)$ (rather than the other possibility $(a + b\bar{\omega})$). Working mod $a + b\omega$, replace $\chi_3(x)$ by $x^{(p-1)/3}$ and $\chi_3(1 - x)$ by $(1 - x)^{(p-1)/3}$ in the definition of $J(\chi_3, \chi_3)$, and use the fact

that $\Sigma x^j = 0$ unless $p - 1$ divides j. **21.** (a) Use Problem 6 and the Hasse–Davenport relation. **22.** (a) $a \neq 0$ and char $K \neq 3$. (b) Use: the map $x \mapsto x^3$ is one-to-one from K to itself, and so $x^3 = u$ always has exactly one solution. (c) There is one point at infinity. For any fixed y, the number of x such that $x^3 = y^2 + ay$ is $1 + \chi_3(y^2 + ay) + \bar{\chi}_3(y^2 + ay)$. Sum this over $y \in K$, making the change of variable $y = ax$, so that $\chi_3(y^2 + ay) = \bar{\chi}_3(a)\chi_3(x - x^2) = \bar{\chi}_3(a)\chi_3(x)\chi_3(1 - x)$ (recall that $-1 = 1$ in K). (d) Compute the case $r = 1$ directly; then use the Hasse–Davenport relation. $Z(C/\mathbb{F}_2; T) = (1 + 2T^2)/(1 - T)(1 - 2T)$. (e) Completing the square in the equation $y^2 + y = x^3$ and substituting $y' = y + \frac{1}{2}$, $x' = x$ gives $y'^2 = x'^3 + \frac{1}{4}$, i.e., $(8y')^2 = (4x')^3 + 16$. Then set $y = 8y'$, $x = 4x'$. **23.** (c) The following tables give the number of points in factored form. The type of the group follows from Problem 10 of §II.1 unless that number is divisible by the square of a prime $l \equiv 1 \pmod 4$. Those cases are marked with an asterisk and discussed below.

r	N_r^{res}	N_r^{nr}
1	2^3	2^2
3	$2^3 \cdot 13$	$2^2 \cdot 37$
5	$2^3 \cdot 401$	$2^2 \cdot 761$
7	$2^3 \cdot 29 \cdot 337$	$2^2 \cdot 29 \cdot 673$

r	N_r
2	2^5
4	$2^7 \cdot 5$
6	$2^5 \cdot 13 \cdot 37$
8	$2^9 \cdot 3^2 \cdot 5 \cdot 17$

r	N_r
10	$2^5 \cdot 401 \cdot 761$
12*	$2^7 \cdot 5 \cdot 13^2 \cdot 37 \cdot 61$
14*	$2^5 \cdot 29^2 \cdot 337 \cdot 673$

(d)

r	N_r^{res}	N_r^{nr}
1	2^3	$2^2 \cdot 5$
3	$2^3 \cdot 277$	$2^2 \cdot 5 \cdot 109$
5	$2^3 \cdot 101 \cdot 461$	$2^2 \cdot 5^2 \cdot 3701$*

r	N_r
2	$2^5 \cdot 5$
4*	$2^7 \cdot 3^2 \cdot 5^2$
6	$2^5 \cdot 5 \cdot 109 \cdot 277$

r	N_r
8*	$2^9 \cdot 3^2 \cdot 5^2 \cdot 73 \cdot 97$
10*	$2^5 \cdot 5^2 \cdot 101 \cdot 461 \cdot 3701$

To handle the asterisked cases, suppose that $E_n(\mathbb{F}_q)$ contains exactly l points of order l: jP, $0 \leq j < l$; and that $E_n(\mathbb{F}_{q^r})$ contains exactly l^2 points of order l^2 of the form jQ, $0 \leq j < l^2$. Let $\mathbb{F}_{q^{r'}} \subset \mathbb{F}_{q^r}$ be the extension generated by the coordinates of Q. Following Problem 12 of §II.1, use the map $\sigma \mapsto \sigma(Q) - Q$ on $\mathrm{Gal}(\mathbb{F}_{q^{r'}}/\mathbb{F}_q)$ to show that $r' = l$. Use this to show that the l part of the type is (l, l) in all cases except N_5^{nr} and N_{10} in part (d). In those two cases, show the type is (l^2) as follows. If there were l^2 $\mathbb{F}_{q^r}$-points of order l, let $\mathbb{F}_{q^{r'}}$ be the extension of $\mathbb{F}_q$ they generate. By suitably choosing a basis $\{P, Q\}$ for the 2-dimensional $\mathbb{F}_l$-vector space of points of order l, get an injection of $\mathrm{Gal}(\mathbb{F}_{q^{r'}}/\mathbb{F}_q)$ into matrices of the form $\left(\begin{smallmatrix}1 & 0\\ 0 & *\end{smallmatrix}\right) \approx \mathbb{F}^*$. Thus, $r' | l - 1$.

§II.3

4. By (3.11), we have $g((\frac{}{p})) = \Sigma_{a,b=0,1,\ldots,p-1} (\frac{a^2+b^2}{p}) e^{2\pi i a/p} = p - 1 + \Sigma_{a=1}^{p-1} \Sigma_{b=0}^{p-1} (\frac{a^2+b^2}{p}) e^{2\pi i a/p}$. In the inner sum replace b by ab to obtain $g((\frac{}{p})) = p - 1 + (\Sigma_{a=1}^{p-1} e^{2\pi i a/p})(\Sigma_{b=0}^{p-1} (\frac{1+b^2}{p})) = p - 1 - \Sigma_{b=0}^{p-1} (\frac{1+b^2}{p}) = 2p - 1 - \Sigma_{b=0}^{p-1} (1 + (\frac{1+b^2}{p})) = 2p - 1 - \#\{a, b \in \mathbb{F}_p | a^2 = 1 + b^2\} = 2p - 1 - \#\{x, y \in \mathbb{F}_p | xy = 1\}$ (after the change of variables $x = a + b$, $y = a - b$; $a = (x + y)/2$, $b = (x - y)/2$). Hence, $g(\frac{}{p})) = 2p - 1 - (p - 1) = p$.

§II.4

2. (b) Note that $L(\chi, s) = \Sigma_{b \in G} \chi(b) f_s(b)$, and use part (a) together with Problem 9(a), (b) in §II.2. (c) $L(\chi, 1) = -\frac{1}{N} g(\chi) \Sigma_{a \in G} \bar{\chi}(a) \log(1 - \xi^{-a})$. (d) $L(\chi, 2) =$

$\frac{1}{N}g(\chi)\sum_{a\in G}\bar{\chi}(a)l(\xi^{-a})$. **3.** (a) $t^{-1/2}e^{2\pi iay-\pi y^2/t}$. (e) $\pi^{-s/2}\Gamma(\frac{s}{2})N^sL(\chi, s) = \pi^{-(1-s)/2}\Gamma(\frac{1-s}{2})g(\chi)L(\bar{\chi}, 1-s)$. (f) Set $s=\frac{1}{2}$ in part (e). (h) $L'(\chi, -2k) = \frac{1}{2}(-1)^k k!\,\pi^{-2k-1/2}N^{2k}g(\chi)\Gamma(k+\frac{1}{2})L(\bar{\chi}, 2k+1)$. **5.** (a) Integrate by parts in the definition of the Fourier transform of $f'(x)$. (b) $-iyt^{-3/2}e^{2\pi iay-\pi y^2/t}$. (e) $\pi^{-(s+1)/2}\Gamma(\frac{s+1}{2})N^sL(\chi, s) = -ig(\chi)\pi^{-(2-s)/2}\Gamma(\frac{2-s}{2})L(\bar{\chi}, 1-s)$. (h) For $\chi(n) = (\frac{n}{3})$ we obtain: $L'(\chi, -1) = \frac{3i}{4\pi}(l(e^{-2\pi i/3}) - l(e^{2\pi i/3}))$, where $l(x)$ is the function in Problem 2(d). **7.** Use part (a)(iii) of Problem 4. **8.** (a) Add the expression for $l(a, 1-s) + l(1-a, 1-s)$ in Problem 3(c) to the expression for $l(a, 1-s) - l(1-a, 1-s)$ in Problem 5(d) to obtain:

$$l(a, 1-s) = \frac{\pi^{-s+1/2}\Gamma(s/2)}{2\Gamma((1-s)/2)}(\zeta(a, s) + \zeta(1-a, s)) + \frac{i\pi^{-s+1/2}\Gamma((s+1)/2)}{2\Gamma(1-s/2)}(\zeta(a, s) - \zeta(1-a, s)).$$

§II.5

2. (a) $\theta(t) = \frac{1}{t}\theta(1/t)$. (b) Let $\phi(s) = \int_1^\infty t^s(\theta(t)-1)\frac{dt}{t} + \int_0^1 t^s(\theta(t)-\frac{1}{t})\frac{dt}{t}$. As in §II.4, show that $\phi(s)$ is entire, and for $\mathrm{Re}\, s > 1$ is equal to $\int_0^\infty t^s(\theta(t)-1)\frac{dt}{t} + \frac{1}{s} + \frac{1}{1-s} = \frac{1}{s} + \frac{1}{1-s} + \pi^{-s}\Gamma(s)\sum_{0\neq m\in\mathbb{Z}^2}|m|^{-2s} = \frac{1}{s} + \frac{1}{1-s} + \pi^{-s}\Gamma(s)4\zeta_K(s)$. By substituting $1/t$ for t in the integrals for $\phi(s)$, show that $\phi(1-s) = \phi(s)$, and hence $\pi^{-s}\Gamma(s)\zeta_K(s) = \pi^{s-1}\Gamma(1-s)\zeta_K(1-s)$. Finally, show that $4\zeta_K(s) = \frac{\pi^s}{\Gamma(s)}(\phi(s) - \frac{1}{s} - \frac{1}{1-s})$ is analytic except at $s=1$, and that $\lim_{s\to1}(s-1)4\zeta_K(s) = \lim_{s\to1}\pi^s/\Gamma(s) = \pi$. **3.** Since the Fourier transform of $g(x) = e^{2\pi iv\cdot x}e^{-\pi t|x+u|^2}$ is $\hat{g}(y) = \frac{1}{t}e^{2\pi iu\cdot(y-v)}e^{-(\pi/t)|y-v|^2}$, one obtains: $\theta_u^v(t) = \frac{1}{t}e^{-2\pi iu\cdot v}\theta_{-v}^u(\frac{1}{t})$. **4.** (a) $(2\pi iw\cdot y)^k\hat{f}(y)$. (b) $(\frac{w\cdot y}{it})^k\frac{1}{t}e^{2\pi iu\cdot y}e^{-(\pi/t)|y|^2}$. (c) $\theta_{u,k}(t) = i^{-k}t^{-k-1}\theta^{u,k}(\frac{1}{t})$. (d) Let $a+bi$ run through a set of coset representatives modulo $I = (n')$, and write x in the sum in the form $(a+bi) + n'(m_1+m_2i)$, so that the sum becomes $n'^k|n'|^{-2s}\sum_{a+bi}\chi(a+bi)\sum_{m\in\mathbb{Z}^2}\frac{((m+u)\cdot w)^k}{|m+u|^{2s}}$ (where $u_1+u_2i = (a+bi)/n'$). The inner sum is essentially the Mellin transform of $\theta_{u,k}(t)$. The functional equation in part (c) will then give a linear combination of terms of the form $\int_0^\infty t^{k+1-s}\theta^{u,k}(t)\frac{dt}{t}$, and this linear combination can be expressed in terms of a Gauss sum and our original sum with s replaced by $k+1-s$. (e) Suppose the (σ_0, σ_1) for our Hecke character $\tilde{\chi}$ is (k_1, k_2) (this pair of integers is called the "infinity type" of $\tilde{\chi}$). Consider $\tilde{\chi}$ as a function on elements as well as ideals of $\mathbb{Z}[i]$ by defining $\tilde{\chi}(x) = \tilde{\chi}((x))$. Let $k = |k_1 - k_2|$. If $k_1 > k_2$, then the mapping $x \mapsto \chi(x) \underset{\text{def}}{=} \tilde{\chi}(x)/x^k(\mathbb{N}x)^{k_2}$ is easily seen to be a character of $(\mathbb{Z}[i]/\mathfrak{f})^*$; if $k_2 > k_1$, then $\chi(x) \underset{\text{def}}{=} \tilde{\chi}(\bar{x})/x^k(\mathbb{N}x)^{k_1}$ is a character of $(\mathbb{Z}[i]/\bar{\mathfrak{f}})^*$. In, for example, the case $k_1 > k_2$, the Hecke L-series is then $\frac{1}{4}\sum_{0\neq x\in\mathbb{Z}[i]}\tilde{\chi}(x)(\mathbb{N}x)^{-s} = \frac{1}{4}\sum x^k\chi(x)(\mathbb{N}x)^{k_2-s}$. **5.** (a) Make the change of variables $x' = Mx$ to obtain $\hat{g}(y) = \int_{\mathbb{R}^n}e^{-2\pi iM^{-1}x'\cdot y}f(x')\frac{dx'}{|\det M|} = \frac{1}{|\det M|}\hat{f}(M^*y)$, since $M^{-1}x'\cdot y = x'\cdot M^*y$. (b) $L' = M^*\mathbb{Z}^n$; let $g(x) = f(Mx)$. Then $\sum_{x\in L}f(x) = \sum_{m\in\mathbb{Z}^n}g(m) = \sum_{m\in\mathbb{Z}^n}\hat{g}(m) = \frac{1}{|\det M|}\sum_{y\in L'}\hat{f}(y)$ by part (a).

6. (a) $M^* = \frac{2}{\sqrt{3}}\begin{pmatrix}\sqrt{3}/2 & 0\\ 1/2 & 1\end{pmatrix}$; considered in $\mathbb{C}$, $L' = \frac{2i}{\sqrt{3}}\mathbb{Z}[\omega]$. Note that $\mathrm{Tr}\, x\bar{y} = 2x\cdot y$, where on the left x and y are considered as elements of $\mathbb{C}$, and on the right as elements of $\mathbb{R}^2$. (b) In Problem 5(b), let $f(x) = e^{-\pi t|x|^2}$, $g(x) = f(Mx)$ with M as in part (a); then apply Poisson summation. (c) Let $\theta(t)$ be the sum on the left in part (b), and let $\phi(s) = \int_{2/\sqrt{3}}^\infty t^s(\theta(t)-1)\frac{dt}{t} + \int_0^{2/\sqrt{3}} t^s(\theta(t) - \frac{2}{t\sqrt{3}})\frac{dt}{t}$. For $\mathrm{Re}\, s > 1$ show that

$\phi(s) = (2/\sqrt{3})^s(\frac{1}{s} + \frac{1}{1-s}) + \pi^{-s}\Gamma(s)6\zeta_K(s)$. The residue at $s = 1$ is $\pi/3\sqrt{3}$. Finally, replacing t by $4/3t$ in the integrals for $\phi(s)$ leads to the relation: $\phi(s) = (2/\sqrt{3})^{2s-1}\phi(1-s)$, and this leads to $\Lambda(s) = \Lambda(1-s)$ for $\Lambda(s) \underset{\text{def}}{=} (\sqrt{3}/2\pi)^s\Gamma(s)\zeta_K(s)$. (d) Use the Euler product form $\zeta_K(s) = \Pi_P(1 - (\mathbb{N}P)^{-s})^{-1}$. For a prime ideal P of norm $p \equiv 1 \pmod 3$, the contribution from P and $\bar{P}$ is $(1-p^{-s})^{-2}$; for $P = (p)$, where $p \equiv 2 \pmod 3$, the contribution is $(1-p^{-2s})^{-1} = (1-p^{-s})^{-1}(1-\chi(p)p^{-s})^{-1}$; and for $P = (\sqrt{-3})$ the contribution is $(1-3^{-s})^{-1}(1-\chi(3)3^{-s})^{-1}$ (since $\chi(3) = 0$). So the Euler product is the Euler product for $\zeta(s)$ times the Euler product for $L(\chi, s)$. (e) Multiplying the functional equations for $\zeta(s)$ and for $L(\chi, s)$ gives $\Lambda(s) = \Lambda(1-s)$ for $\Lambda(s) = \pi^{-s/2}\Gamma(\frac{s}{2})\zeta(s)(3/\pi)^{s/2}\Gamma((s+1)/2)L(\chi, s) = (\sqrt{3}/\pi)^s\zeta_K(s)\Gamma(\frac{s}{2})\Gamma((s+1)/2) = \text{const}\cdot(\sqrt{3}/2\pi)^s\Gamma(s)\zeta_K(s)$ by (4.4). **7.** See Problems 20–22 in §II.2; (d) $g(\chi, \psi) = 3$. **8.** (a) By part (1) of Proposition 9, $\hat{g}(y)$ equals $e^{2\pi i u\cdot y}$ times the Fourier transform of $(x\cdot w)e^{-\pi t|x\cdot w|^2}$; note that $x\cdot w = Mx\cdot(1, i)$ with M as in Problem 6(a); then proceed as in Problem 6(b) to obtain $\hat{g}(y) = \frac{4}{3t^2}y\cdot(-\omega, 1)e^{2\pi i u\cdot y}e^{-(4\pi/3t)|y\cdot(-\omega,1)|^2}$. (b) Use Problem 7(c) to obtain $\phi(s) = 3^{2s-1}\pi^{-s}\Gamma(s)6L(E, s)$. (d) Replacing t by $\frac{4}{3t}$ in the integrals in $\phi(s)$, one obtains $\phi(s) = (\frac{4}{3})^{s-1}\Sigma\chi(a + b\omega)\int_0^\infty t^{2-s}\theta^u(t)\frac{dt}{t}$, where the summation is over $0 \le a, b < 3$; $u = (a/3, b/3)$; and $\theta^u(t) \underset{\text{def}}{=} \Sigma_{m\in\mathbb{Z}^2} m\cdot(-\omega, 1)e^{2\pi i u\cdot m}e^{-\pi t|m\cdot(-\omega,1)|^2}$. Then for Re $(2-s) > 3/2$ use (4.6), Problem 2 of §II.2 (note that $u\cdot m = \frac{1}{3}\text{Tr}((a+b\omega)(-\bar{\omega}m_1 + m_2)/i\sqrt{3})$), and the evaluation of the Gauss sum in Problem 7(d). The result is: $\phi(s) = (4/3)^{s-1}3\pi^{s-2}\Gamma(2-s)6L(E, 2-s)$. Equating this with the expression in part (b) and collecting terms gives the desired result.

§II.6

2. The function $f(s) = \Lambda(1+s)$ is even in the first case (so that its Taylor expansion at $s = 0$ has only even powers of s) and odd in the second case. **3.** (c) Use: $b_{p^e} = \alpha_p^e + \alpha_p^{e-1}\bar{\alpha}_p + \cdots + \bar{\alpha}_p^e$, and $2a_p = \alpha_p + \bar{\alpha}_p$.

5–6.

n	first few nonzero $b_{m,n}$	$L(E_n, 1)$	remainder estimate
2	$b_1 = 1, b_5 = 2, b_9 = -3$	0.92707	$\|R_{13}\| \le 0.00027$
3	$b_1 = 1, b_5 = 2, b_{13} = 6, b_{17} = -2$	1.5138	$\|R_{25}\| \le 0.00123$
10	$b_1 = 1, b_9 = -3, b_{13} = 6, b_{17} = -2$	1.65	$\|R_{29}\| \le 0.289$

7. You want $|R_{M+1}|$ to be less than $c/2$, i.e., $c/2 > 4(1 - e^{-\pi/\sqrt{N'}})^{-1}e^{-\pi(M+1)/\sqrt{N'}}$; for large n the right side is asympotic to $\frac{4}{\pi}\sqrt{N'}e^{-\pi M/\sqrt{N'}}$, so choose $M > \frac{1}{\pi}\sqrt{N'}\log(8\sqrt{N'}/\pi c) \approx \frac{1}{\pi}\sqrt{N'}\log n$, i.e. $(2n\sqrt{2}/\pi)\log n$ for n odd, $(2n/\pi)\log n$ for n even. **8.** (a) Use Problem 3 to find the $b_{m,n}$. (b) See Problem 7 of §I.1.

§III.1

1. $T \in \Gamma_1(N) \subset \Gamma_0(N)$, but $STS^{-1} \notin \Gamma_0(N)$; hence, neither $\Gamma_1(N)$ nor $\Gamma_0(N)$ is normal in Γ. **2.** (a) Given $\bar{A} \in SL_2(\mathbb{Z}/N\mathbb{Z})$, let A be any matrix with $\bar{A} = A \bmod N$. Find $B, C \in SL_2(\mathbb{Z})$ such that BAC is diagonal: $BAC = (\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix})$, where $ad = \det A \equiv 1 \pmod N$. It suffices to find $\tilde{A} = (\begin{smallmatrix} a+xN & zN \\ N & d \end{smallmatrix})$ with determinant 1, since then $B^{-1}\tilde{A}C^{-1} \in SL_2(\mathbb{Z})$, and $B^{-1}\tilde{A}C^{-1} \equiv B^{-1}(\begin{smallmatrix} a & 0 \\ 0 & d \end{smallmatrix})C^{-1} = A \equiv \bar{A} \pmod N$. To find $\tilde{A}$ so that $1 = \det\tilde{A} = 1 + ((ad-1)/N + xd)N - zN^2$, first find x so that $xd \equiv -(ad-1)/N \pmod N$, and then choose z so as to make $\det\tilde{A} = 1$.

3. (a) $(q^2-1)(q^2-q)$; (b) $q(q^2-1)$. **4.** (b) Since the kernel in part (a) has $p^{4(e-1)}$ elements and $\#GL_2(\mathbb{Z}/p\mathbb{Z}) = (p^2-1)(p^2-p)$, it follows that $\#GL_2(\mathbb{Z}/p^e\mathbb{Z}) = p^{4e-3}(p^2-1)(p-1)$. (c) Divide the answer in (b) by $\phi(p^e)$ to get $p^{3e-2}(p^2-1)$
5. Use the Chinese remainder theorem. **6.** $N^3\Pi_{p|N}(1-p^{-2})$.
7. $N^3\Pi_{p|N}(1-p^{-2})$; N; $N\Pi_{p|N}(1-p^{-1})$; $N^2\Pi_{p|N}(1-p^{-1})$; $N\Pi_{p|N}(1+p^{-1})$
8. The image of $\Gamma(N)$ under conjugation by $\left(\begin{smallmatrix}0 & -1\\ N & 0\end{smallmatrix}\right)$ is the subgroup of $\Gamma_0(N^2)$ consisting of matrices whose upper-left entry is $\equiv 1 \pmod N$. **10.** and 14(b). See Figures A.1–A.2. **12.** Besides Γ and $\Gamma(2)$, there are four, namely, the preimages of the three subgroups of order 2 in S_3 and the one subgroup of order 3 under the map $\Gamma \to SL_2(\mathbb{Z}/2\mathbb{Z}) \approx S_3$. (i) $\Gamma_0(2) = \{\left(\begin{smallmatrix}* & *\\ 0 & *\end{smallmatrix}\right) \bmod 2\}$, $F_0(2) =$ the right half of $F(2)$; (ii) $\Gamma^0(2) = \{\left(\begin{smallmatrix}* & 0\\ * & *\end{smallmatrix}\right) \bmod 2\}$, $F^0(2) = F \cup T^{-1}F \cup SF$; (iii) $\mathfrak{G}(2) \underset{\text{def}}{=} \{\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) \equiv I \text{ or } S \bmod 2\}$, fundamental domain $= F \cup T^{-1}F \cup T^{-1}SF$; (iv) $\{\left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) \equiv I, ST \text{ or } (ST)^2 \bmod 2\}$, fundamental domain $= F \cup T^{-1}F$. **13.** (a) This can be proved using the fundamental domain, as in the proof of Proposition 4 in the text. Here is another method. Let G denote the subgroup of $\mathfrak{G}(2)$ generated by $\pm S$, T^2. Clearly $G \subset \mathfrak{G}(2)$. Conversely, write $g \in \mathfrak{G}(2)$ as a word of the form $\pm S^{a_1}T^{b_1}ST^{b_2}\cdots ST^{b_l}$, where $a_1 = 0$ or 1 and $b_j \neq 0, j = 1, \ldots, l-1$. Use induction on l to show that $g \in G$. We work mod $\pm I$, so that $S^2 = (ST)^3 = 1$. Without loss of generality we may suppose $a_1 = 0$, $b_l \neq 0$, since we can always multiply g on the left or right by S without affecting whether $g \in G$. For the same reason we may suppose that $b_1 = b_l = 1$, since $T^2 \in G$. Note that $l = 1$ or 2 is impossible, since $T, TST \notin \mathfrak{G}(2)$. If $l > 2$, write $g = TST^{b_2}\cdots$. Since $(STS)(TST) = 1$, we have $TST = (STS)^{-1} = ST^{-1}S$, and so $T^2Sg = TST^{b_2-1}S\cdots$, which is just like g but with b_2 replaced by $b_2 - 1$. If $b_2 > 0$, use induction on b_2 to finish the proof. If $b_2 < 0$, write $ST^{-2}g = ST^{-1}ST^{b_2}\cdots = TST^{b_2+1}\cdots$, and again use induction on $|b_2|$. (b) Use: $\Gamma^0(2) = T\mathfrak{G}(2)T^{-1}$, $\Gamma_0(2) = ST\mathfrak{G}(2)(ST)^{-1}$. (c) Let G be the subgroup of $\bar{\Gamma}$ generated by T^2 and $ST^{-2}S$. Since $G \subset \bar{\Gamma}(2)$, it suffices to show that $[\bar{\Gamma}: G] = 6$, e.g., that any "word" in S and T can be multiplied on the left by elements of G to obtain one of the elements α_j used in the text for coset representatives. Use $S^2 = (ST)^3 = 1$ and induction as in part (a). (d) Use part (c) and the isomorphism in Problem 8. **14.** (c) $F_0(p)$ is bounded by the vertical lines above $\frac{1}{2}$ and $-\frac{1}{2}$; arcs of circles with diameter $[0, 2]$ and $[-2/(2p-1), 0]$; and arcs of circles with diameter $[-1/(k-2), -1/k]$,

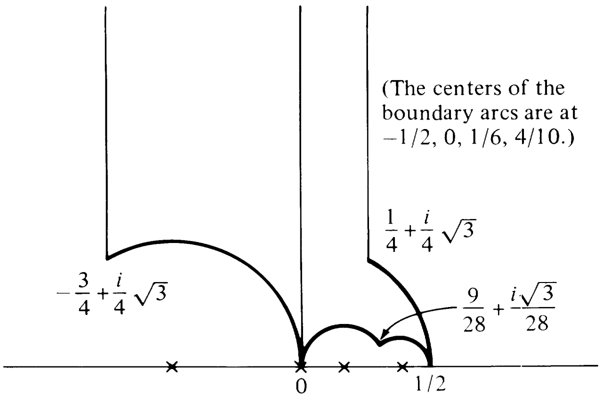


Figure A.1

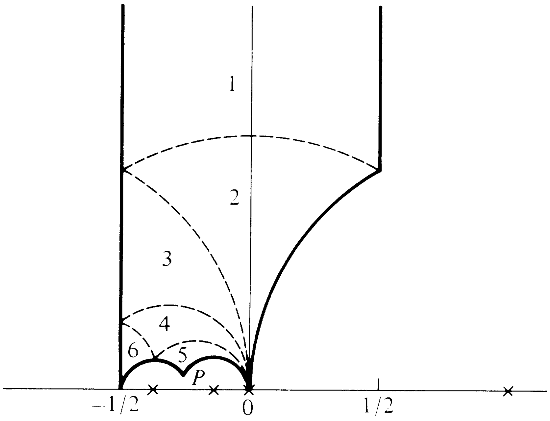


Figure A.2. Two possible fundamental domains for $\Gamma_0(4)$:

I. Let $F(2)$ be the fundamental domain for $\Gamma(2)$ in the text. Then $\alpha F(2)$, where $\alpha = \left(\begin{smallmatrix} 0 & -1 \\ 2 & 0 \end{smallmatrix}\right)$, is a fundamental domain for $\Gamma_0(4) = \alpha\Gamma(2)\alpha^{-1}$ (see Fig. A-1).

II. $\bigcup_{j=1}^{6} \alpha_j^{-1} F$ is a fundamental domain for $\Gamma_0(4)$, where F is the fundamental domain for $\Gamma = SL_2(\mathbb{Z})$ given in the text, and α_j are coset representatives for Γ modulo $\Gamma_0(4)$. Here we have taken: $\alpha_1 = 1$, $\alpha_2 = S$, $\alpha_3 = T^{-1}S$, $\alpha_4 = T^{-2}S$, $\alpha_5 = T^{-3}S$, $\alpha_6 = ST^{-2}S$. (In Fig. A.2, the number j labels $\alpha_j^{-1}F$; the solid boundary arcs are centered at $1, -\frac{1}{7}, -\frac{3}{8}$; the dotted arcs are centered at $0, -1, -\frac{1}{3}, -\frac{2}{3}, -\frac{1}{5}$; the point P is $(-7 + i\sqrt{3})/26$.)

$k = 3, \ldots, p$. For example, $F_0(3)$ is the union of the regions marked 1, 2, 3 and 4 in Figure A.2. **15.** See Problems 12 and 14(c) for counterexamples. **16.** (b) (i) 2; (ii) 2; (iii) 2; (iv) 1.

§III.2

1. In the sum for G_k, group together indices m, n with given g.c.d. **2.** Substitute $z = i$ in Proposition 7 to obtain: $E_2(i) = 3/\pi$. **3.** (a) Since both sides of each equality are in a one-dimensional $M_k(\Gamma)$, by Proposition 9(c), it suffices to check equality of constant terms. (b) Equate coefficients of q^n on both sides of the equalities in part (a) to obtain: $\sigma_7(n) = \sigma_3(n) + 120\sum_{j=1}^{n-1} \sigma_3(j)\sigma_3(n-j)$; $11\sigma_9(n) = -10\sigma_3(n) + 21\sigma_5(n) + 5040\sum_{j=1}^{n-1} \sigma_3(j)\sigma_5(n-j)$; $\sigma_{13}(n) = 21\sigma_5(n) - 20\sigma_7(n) + 10080\sum_{j=1}^{n-1} \sigma_5(j)\sigma_7(n-j)$. **4.** (a) Since the left side is in $S_{12}(\Gamma)$, by Proposition 9(d) it suffices to check equality of coefficients of q. (b) $\tau(n) = \frac{65}{756}\sigma_{11}(n) + \frac{691}{756}\sigma_5(n) - \frac{691}{3}\sum_{j=1}^{n-1} \sigma_5(j)\sigma_5(n-j)$. (c) Consider part (b) modulo 691. **5.** Let $F(X, Y)$ be a homogeneous irreducible polynomial satisfied by $X = E_4$, $Y = E_6$. Substituting $z = i$ in $E_4(z)$, $E_6(z)$ leads to a contradiction, since $E_6(i) = 0$, $E_4(i) \neq 0$. **6.** (b) Use part (a) with $x = e^{-2\pi} = e^{2\pi i(i)}$, along with the relation $0 = \frac{1}{2(a+1)}B_{a+1}E_{a+1}(i) = \frac{1}{2(a+1)}B_{a+1} - \sum_{n=1}^{\infty} \sigma_a(n)q^n$ with $q = e^{2\pi i(i)}$. **7.** Use Proposition 7 and the derivative of the identity $f(-1/z) = z^k f(z)$. **8.** (a) By Problem 7, the right sides are in $M_6(\Gamma)$ and $M_8(\Gamma)$, respectively. Now proceed as in Problem 3(a). (b) $21\sigma_5(n) = 10(3n-1)\sigma_3(n) + \sigma_1(n) + 240\sum_{j=1}^{n-1} \sigma_1(j)\sigma_3(n-j)$; $20\sigma_7(n) = (42n - 21)\sigma_5(n) - \sigma_1(n) + 504\sum_{j=1}^{n-1} \sigma_1(j)\sigma_5(n-j)$. **9.** (a) Use the fact

that $n^2 \equiv 1 \pmod{24}$ if n is prime to 12. (b) Use Proposition 9(d) to show that their 24-th powers are equal. (c) $\Pi_{n=1}^{\infty}(1-q^n) = \Sigma_{n\equiv \pm 1 \bmod 12} q^{(n^2-1)/24} - \Sigma_{n\equiv \pm 5 \bmod 12} q^{(n^2-1)/24}$. **10.** $j = 256(\lambda^{-2}+1+\lambda^2)^3/(\lambda+\lambda^{-1})^2$; $j = 1728$ if $\lambda = 1$; if $\lambda = a/b$ in lowest terms (with a and b positive) and $j \in \mathbb{Z}$, that means that $a^4b^4(a^2+b^2)^2$ divides $256(a^4+a^2b^2+b^4)^3$, but since a, b and a^2+b^2 each have no common factor with $a^4+a^2b^2+b^4$, that means that $a^2b^2(a^2+b^2)$ divides 16, and now it's easy to eliminate all possibilities except $a = b = 1$.

§III.3

4. (a) Since $\Gamma_0(N) = \pm\Gamma_1(N)$ in those cases, there is no difference between $\Gamma_0(N)$-equivalence and $\Gamma_1(N)$-equivalence. (b) $-1/2$ is an irregular cusp for $\Gamma_0(4)$.

5.

Group	$\Gamma_0(p)$		$\Gamma_0(p^2)$			$\Gamma(2)$		
Cusp	∞	0	∞	0	$-1/kp,\ k=1,\ldots,p-1$	∞	0	-1
Index	1	p	1	p^2	1	2	2	2

7. (a) See the proof of Proposition 18; replace a by X in (3.11). **8.** (a) Replacing z by $z+1/2$ in $\Sigma e^{2\pi i z n^2}$ gives $\Sigma(-1)^n e^{2\pi i z n^2}$; meanwhile, the right-hand side is $2\Sigma e^{2\pi i z(2n)^2}$. (b) works the same way. (c) On the left side, the constant term is clearly zero; the coefficient of q^n for $p \nmid n$ is $(-2k/B_k)\sigma_{k-1}(n)$; if $p|n$ but $p^2 \nmid n$, then the coefficient is $-2k/B_k$ times $\sigma_{k-1}(n) - (1+p^{k-1})\sigma_{k-1}(n/p) = 0$; if $n = p^m n_0$ with $m > 1$ and $p \nmid n_0$, then the coefficient of q^n is
$-2k/B_k$ times $\sigma_{k-1}(p^m n_0) - (1+p^{k-1})\sigma_{k-1}(p^{m-1}n_0) + p^{k-1}\sigma_{k-1}(p^{m-2}n_0) =$
$\sigma_{k-1}(n_0)(\sigma_{k-1}(p^m) - (1+p^{k-1})\sigma_{k-1}(p^{m-1}) + p^{k-1}\sigma_{k-1}(p^{m-2})) = 0$.
(d) Use parts (b) and (c) with $p = k = 2$. (e) Rearrange the infinite product of $(1-e^{2\pi i(z+1/2)n}) = (1-(-1)^n q^n)$ by writing $1+q^n = (1-q^{2n})/(1-q^n)$, getting $\Pi_{n\text{ even}} \cdot \Pi_{n\text{ twice an odd}}/\Pi_{n\text{ odd}}$ (where Π denotes $\Pi(1-q^n)$). But this equals $\Pi^2_{n\text{ even}} \cdot \Pi_{n\text{ even}}/\Pi_{\text{all }n} \cdot \Pi_{n\text{ twice an even}} = \Pi^3_{2|n}/\Pi_n \cdot \Pi_{4|n}$. **9.** See the proof of Proposition 30. **10.** (a) Since $\eta^8(z)\eta^8(2z) \in S_8(\Gamma_0(2))$ by Proposition 20, to show invariance of $\eta^8(4z)/\eta^4(2z)$ under $[\gamma]_2$ for $\gamma \in \Gamma_0(4)$ it suffices to show invariance of $\eta^8(4z)\eta^8(z)\eta^4(2z)$ under $[\gamma]_{10}$ for $\gamma \in \Gamma_0(4)$; now use Problem 9 to show this. At the cusp ∞, we see that $\eta^8(4z)/\eta^4(2z) = q\Pi(1-q^{4n})^8/(1-q^{2n})^4$ has a first order zero; to find $\eta(0)$ we apply $[S]_2$: $z^{-2}\eta^8(-4/z)/\eta^4(-2/z) =$
$z^{-2}(z/4)^4\eta^8(z/4)/(-(z/2)^2\eta^4(z/2)) = -\frac{1}{64}\Pi(1-q_4^n)^8/(1-q_4^{2n})^4$, which approaches $-\frac{1}{64}$ as $z \to \infty$. To find the value at the cusp $-\frac{1}{2}$, which is equivalent to the cusp $\frac{1}{2} = T(-1/2)$, we can apply $[(\begin{smallmatrix}1 & 0\\ 2 & 1\end{smallmatrix})]_2$, as follows:

$$(2z+1)^{-2}\frac{\eta^8\left(\dfrac{4z}{2z+1}\right)}{\eta^4\left(\dfrac{2z}{2z+1}\right)} = (2z+1)^{-2}\frac{\left(\dfrac{2z+1}{4z}\right)^4\eta^8\left(-\dfrac{1}{4z}-\dfrac{1}{2}\right)}{-\left(\dfrac{2z+1}{2z}\right)^2\eta^4\left(-\dfrac{1}{2z}-1\right)}$$

$$= -\frac{1}{64z^2}e^{-2\pi i/6}\frac{\eta^{24}\left(\dfrac{-1}{2z}\right)}{\eta^8\left(\dfrac{-1}{4z}\right)\eta^8\left(-\dfrac{1}{z}\right)}\frac{1}{e^{-2\pi i/6}\eta^4\left(\dfrac{-1}{2z}\right)} \quad \text{by Problem 8(e)}$$

$$= \frac{1}{16}\frac{\eta^{20}(2z)}{\eta^8(z)\eta^8(4z)},$$

which approaches $\frac{1}{16}$ as $z \to \infty$.
(b) $E_2(ST^{-a}Sz) = E_2(S(-a - \frac{1}{z})) = (a + \frac{1}{z})^2 E_2(-a - \frac{1}{z}) + \frac{6}{\pi i}(-a - \frac{1}{z})$, but $E_2(-a - \frac{1}{z}) = E_2(-\frac{1}{z})$, so this equals $(a + \frac{1}{z})^2(z^2 E_2(z) + \frac{6z}{\pi i}) + \frac{6}{\pi i}(-a - \frac{1}{z}) = (az + 1)^2 E_2(z) - \frac{6ai}{\pi}(az + 1)$. (c) Obviously $F|[T]_2 = F$. Using Problems 1 and 10(b), we have $-24F(z)|[ST^{-4}S]_2 = E_2(z)|[ST^{-4}S]_2 - 3E_2(2z)|[ST^{-2}S]_2 + 2E_2(4z)|[ST^{-1}S]_2 = E_2(z) + \frac{24}{\pi i}\frac{1}{(4z+1)} - 3(E_2(2z) + \frac{12}{\pi i}\frac{1}{(4z+1)}) + 2(E_2(4z) + \frac{6}{\pi i}\frac{1}{(4z+1)}) = -24F(z)$. We now look at the cusps. First, we clearly have $F(\infty) = 0$. In evaluating $F|[S]_2$, ignore terms which approach zero as $z \to \infty$; then, using Proposition 7, obtain $F(z)|[S]_2 \sim -\frac{1}{24}E_2(z) - 3(\frac{1}{4}E_2(z/2)) + 2(\frac{1}{16}E_2(z/4)) \to -\frac{1}{24}(1 - \frac{3}{4} + \frac{2}{16}) = -\frac{1}{64}$. This is $F(0)$. To find $F(-\frac{1}{2})$, we similarly ignore terms which approach zero: $-24F(z)|[ST^{-2}S]_2 = E_2(z)|[ST^{-2}S]_2 - 3E_2(2z)|[ST^{-1}S]_2 + 2(2z + 1)^{-2}E_2(\frac{4z}{2z+1})$ by Problem 1 above, and, by Proposition 7 and part (b), neglecting terms which approach zero, this becomes $E_2(z) - 3E_2(2z) + 2(2z + 1)^{-2}(-\frac{1}{4z} - \frac{1}{2})^2 E_2(-\frac{1}{4z} - \frac{1}{2}) = E_2(z) - 3E_2(2z) + \frac{1}{8}z^{-2}(-E_2(-\frac{1}{4z}) + 6E_2(-\frac{1}{2z}) - 4E_2(-\frac{1}{z}))$ by Problem 8(d); again applying Proposition 7 and neglecting small terms, we obtain $E_2(z) - 3E_2(2z) + \frac{1}{8}(-16E_2(4z) + 24E_2(2z) - 4E_2(z)) \to 1 - 3 + \frac{1}{8}(-16 + 24 - 4) = -\frac{3}{2}$, so that $F(-\frac{1}{2}) = -\frac{1}{24}(-\frac{3}{2}) = \frac{1}{16}$. (d) Since the only zero of $\eta^8(4z)/\eta^4(2z)$ is a simple zero at ∞, it follows by Proposition 18 that $F(z)/(\eta^8(4z)/\eta^4(2z)) \in M_0(\Gamma_0(4))$ is a constant. To get the identity, write $(1 - q^{4n})/(1 - q^{2n}) = (1 + q^{2n})$. (e) First show that $\Pi_{j=0}^{N-1}\Delta(z + j/N) \in S_{12N}(\Gamma_0(N^2))$, using the same type of argument as in the proof of Proposition 17(b). Then set $f(z) = \Delta(z)^N/\Pi\Delta(z + j/N)$, and take the logarithmic derivative of the equality $f(\gamma z) = f(z)$ for $\gamma \in \Gamma_0(N^2)$. **11.** (a) Prove invariance under $[\gamma]_2$ for $\gamma \in \Gamma_0(4)$ as in Proposition 30. At the cusps, clearly $\Theta^4(\infty) = 1$; at 0 we have $\Theta^4(z)|[S]_2 = z^{-2}\Theta^4(-1/(4z/4)) = z^{-2}(-z^2/4)\Theta^4(z/4)$, which approaches $-\frac{1}{4}$ as $z \to \infty$; at $\frac{1}{2}$ we have $\Theta^4(z)|[ST^{-2}S]_2 = (2z + 1)^{-2}\Theta^4(z/(2z + 1)) = (2z + 1)^{-2}\Theta^4(-1/(4(-1/4z - 1/2))) = -(2z)^{-2}\Theta^4(-1/4z - 1/2)$ by (3.4), but by Problem 8(a) this equals $-(2z)^{-2}(2\Theta(-1/z) - \Theta(-1/4z))^4 = -\frac{1}{4}(2(2i)^{-1/2}\Theta(z/4) - \sqrt{2/i}\,\Theta(z))^4 = (\Theta(z/4) - \Theta(z))^4$, which approaches zero as $z \to \infty$. (b) Use: $\Theta^4(\infty) \neq 0 = F(\infty)$.
(c) One can proceed directly, as in Problem 10(a). Alternately, write it as $\frac{\Delta(2z)}{(\eta(z)\eta(2z))^8}\frac{\eta^4(2z)}{\eta^8(4z)}$, and use Propositions 17 and 20 and Problem 10(a). An even easier method is as follows. Let $f_1(z) = \eta^4(2z)/\eta^8(4z)$, and let $f_2(z) = \eta^{20}(2z)/\eta^8(z)\eta^8(4z)$. In the solution to Problem 10(a), we saw that for $\alpha = ST^{-2}S = -\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}$ we have $f_2 = 16f_1|[\alpha]_2$. Since f_1 is invariant under $[\gamma]_2$ for $\gamma \in \Gamma_0(4)$, it follows that f_2 is invariant under $\alpha^{-1}\Gamma_0(4)\alpha$, which is easily checked to be equal to $\Gamma_0(4)$; finally, since α keeps 0 fixed and interchanges the equivalence classes of cusps ∞ and $-\frac{1}{2}$, we see that $f_2(0) = 16f_1(0) = -\frac{1}{4}$, $f_2(-\frac{1}{2}) = 16f_1(\infty) = 0$, $f_2(\infty) = 16f_1(-\frac{1}{2}) = 1$. (d) Same procedure as 10(d). (e) Use Problem 8(e).
12. Follow the proof of Proposition 19. **13.** Write $(\eta(z)\eta(3z))^6 = \Delta(z)(\eta(3z)/\eta^3(z))^6$, $(\eta(z)\eta(7z))^3 = \Delta(z)(\eta(7z)/\eta^7(z))^3$. **14.** Check the generators T^2 and S, using (3.4) to get $\phi(Sz) = \sqrt{z/i}\,\phi(z)$; to check the cusp $-1 = (TST)^{-1}\infty$, write $\phi^4(z)|[TST]_2 = \Theta^4(z/2)|[\begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}]_2$ (see Problem 1).
15. (a) For $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma_0(N)$ note that $\alpha_N\gamma\alpha_N^{-1} = \begin{pmatrix} d & -c/N \\ -Nb & a \end{pmatrix} \in \Gamma_0(N)$, and so $(f|[\alpha_N]_k)|[\gamma]_k = (f|[\alpha_N\gamma\alpha_N^{-1}]_k)|[\alpha_N]_k = \chi(a)f|[\alpha_N]_k$. But $\chi(a) = \bar{\chi}(d)$.
(b) For $f \in M_k(N, \chi)$ write $f = f^+ + f^-$, where $f^{\pm} = \frac{1}{2}(f \pm i^k f|[\alpha_N]_k)$. (c) By 10(d), $F(z)|[\alpha_4]_2 = -\frac{1}{16}\eta^8(z)/\eta^4(2z) = -\frac{1}{16} + \frac{1}{2}q + \cdots = -\frac{1}{16}\Theta^4 + F$; matrix is $\begin{pmatrix} -1 & -1/16 \\ 0 & 1 \end{pmatrix}$; $M_2^+(4, 1) = \mathbb{C}\frac{1}{8}\Theta^4$, $M_2^-(4, 1) = \mathbb{C}(\frac{4}{3}F - \frac{1}{24}\Theta^4)$.
16. (a) $\Sigma\sigma_{k-1}(n)n^{-s} = \Sigma_{jm=n}j^{k-1}(jm)^{-s} = \Sigma_j j^{k-1-s}\Sigma_m m^{-s} = \zeta(s + 1 - k)\zeta(s)$.

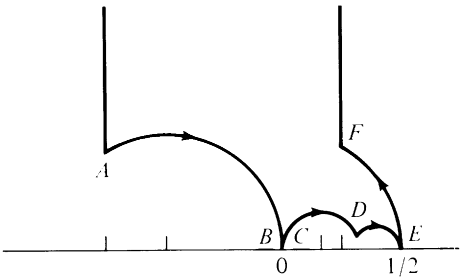


Figure A.3

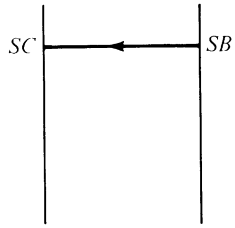


Figure A.4

(b) $L_f(s) = \Pi_p[(1-p^{-s})(1-p^{k-1-s})]^{-1} = \Pi_p(1-\sigma_{k-1}(p)p^{-s}+p^{k-1-2s})^{-1}$.
(c) $L_{f_\chi}(s) = \Pi_p[(1-\chi(p)p^{-s})(1-\chi(p)p^{k-1-s})]^{-1}$
$= \Pi_p(1-\chi(p)\sigma_{k-1}(p)p^{-s}+\chi(p)^2p^{k-1-2s})^{-1}$.

17. (a) Any point Γ-equivalent to i must be $\Gamma_0(4)$-equivalent to one of the points $\alpha_j^{-1}i$, where $\Gamma = \bigcup_{j=1}^{6}\alpha_j\Gamma_0(4)$. In this way, find that i, $(2+i)/5 \in \operatorname{Int} F'$ are Γ-equivalent to i; ω and $(5+i\sqrt{3})/14 \in \operatorname{Int} F'$ are Γ-equivalent to ω; the two $\Gamma_0(4)$-equivalent boundary points $(-1+i)/2$ and $(3+i)/10$ are Γ-equivalent to i; and no boundary points are Γ-equivalent to ω. (b) Follow the proof of Proposition 8, but with slightly more involved computations. Note that the three "corners" $(-3+i\sqrt{3})/4$, $(1+i\sqrt{3})/4$, $(9+i\sqrt{3})/28$ are all $\Gamma_0(4)$-equivalent, and the sum of the angles at the three corners is 360°. To illustrate the elements in the integration around the cusps and along the circular arcs, let us compute $(2\pi i)^{-1}\int f'(z)dz/f(z)$ over the contour ABCDEF pictured in Fig. A-3, where we suppose that $f(z)$ has no zeros or poles on the contour (but may have a zero or pole at the cusp 0), and we ultimately want the limit as $\varepsilon \to 0$. Here BC is a circular arc of radius ε centered at 0. The element $\alpha_1 = \begin{pmatrix}1&0\\4&1\end{pmatrix} \in \Gamma_0(4)$ takes the arc from A to 0 to the arc from D to 0. The image of B is very close to C, and so we can use the integral from D to $\alpha_1 B$ to approximate the integral from D to C. First, we use S to take the arc BC to a horizontal line between $\operatorname{Re} z = -3$ and $\operatorname{Re} z = 1$ (see Fig. A.4). By (2.24), we have

$$\int_B^C \frac{f'(z)}{f(z)}dz = \int_{SB}^{SC}\frac{f'(z)}{f(z)}dz - k\int_B^C \frac{dz}{z}.$$

But the latter integral approaches 0 as $\varepsilon \to 0$ (since the angle of the arc BC approaches zero), while the first integral on the right approaches $-v_0(f)$ (as in the proof of Proposition 8, but we use the map $q_4 = e^{2\pi i z/4}$ into the unit disc). Next, by

(2.24), we have

$$\int_A^B \frac{f'(z)}{f(z)}dz + \int_C^D \frac{f'(z)}{f(z)}dz \sim \int_A^B \frac{f'(z)}{f(z)}dz - \int_{\alpha_1 A}^{\alpha_1 B} \frac{f'(z)}{f(z)}dz = -k\int_A^B \frac{dz}{z+1/4}$$

$$\sim -k\int_A^0 \frac{dz}{z+1/4} = -k\int_{(-2+i\sqrt{3})/4}^{1/4} \frac{dz}{z}.$$

Similarly, using $\alpha_2 = \left(\begin{smallmatrix} 3 & -1 \\ 4 & -1 \end{smallmatrix}\right) \in \Gamma_0(4)$ to take the arc DE to the arc FE, we find that the sum of the integrals over those two arcs is equal to $-k$ times the integral of $\frac{dz}{z}$ from $(2+i\sqrt{3})/28$ to $\frac{1}{4}$. These two integrals are evaluated by taking $\ln z$ between the limits of integration, where the branch is determined by following the contour. As a result, we find that the sum of the two integrals is $-k$ times the logarithm of

$$\frac{1/4}{(-2+i\sqrt{3})/4} \cdot \frac{1/4}{(2+i\sqrt{3})/28} = -1,$$

where, if we keep track of the contour, we see that we must take $\ln(-1) = -\pi i$. Thus, $(2\pi i)^{-1}$ times the sum of these two integrals is equal to $k/2$. For a systematic treatment of formulas for the sum of orders of zero of a modular form for a congruence subgroup, see [Shimura 1971, Chapter 2]. (c) The only zero of Θ^4 is a simple zero at the cusp $-\frac{1}{2}$; the only zero of F is a simple zero at ∞. (d) If $f \in M_2(\Gamma_0(4))$, apply part (b) to the element $f - f(\infty)\Theta^4 - 16f(-\frac{1}{2})\,F \in M_2(\Gamma_0(4))$, which is zero at $-\frac{1}{2}$ and ∞, and hence is the zero function. (e)–(f) See the proof of Proposition 9. (g) Look at the value at each cusp of $f = a\Theta^{12} + b\Theta^8 F + c\Theta^4 F^2 + dF^3 : f(\infty) = a, f(-\frac{1}{2}) = d/16^3$, so $a = d = 0$; then $f(0) = (-\frac{b}{4} - \frac{c}{64})(-\frac{1}{64})(-\frac{1}{4})$, so $c = -16b$. (h) Let $f(z) = \eta^{12}(2z)$. Since $f(z)^2 = \Delta(2z) \in M_{12}(\Gamma_0(2))$, we immediately have vanishing at the cusps. Let $\alpha = \left(\begin{smallmatrix} 1 & 0 \\ -2 & 1 \end{smallmatrix}\right)$. By writing $2\alpha z = -1/(1 - 1/2z)$, show that $f|[\alpha]_6 = -f$, and so $f \notin M_6(\Gamma_0(2))$. However, $\bar{\Gamma}_0(4)$ is generated by T and α^2. Next, since $S_6(\Gamma_0(4)) = \mathbb{C}(\Theta^8 F - 16\Theta^4 F^2)$, to check the last equality in part (h), it suffices to check equality of the coefficient of q. **18.** (b) If $f \in M_1(4, \chi)$, by subtracting off a multiple of Θ^2 we may suppose that $f(\infty) = 0$; then $M_2(4, 1) \ni f^2 = a_1^2 q^2 + \cdots$. But a multiple zero at ∞ would contradict Problem 17(b), unless f is the zero function. (c) First show that, for $k \geq 5$ odd, $S_k(4, \chi) = \Theta^{-2} S_{k+1}(4, 1)$; then, by Problem 17($i$), $\dim S_k(4, \chi^k) = [(k-3)/2]$ (where [ ] is the greatest integer function). (d) Use Proposition 20, Problem 17 (with $\Gamma' = \Gamma_0(2)$, $\alpha = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 1 \end{smallmatrix}\right)$), and part (c).

§III.4

2. Show that $\sqrt{-i(cz+d)}/\sqrt{-i(-cz-d)} = i^{-\operatorname{sgn} d}$ (it suffices to check for $z = 0$), and then divide into four cases, depending on the sign of c and d. **3.** (b)–(c) Use the relation $\dfrac{\Theta(\alpha\beta z)}{\Theta(z)} = \dfrac{\Theta(\alpha\beta z)}{\Theta(\beta z)} \cdot \dfrac{\Theta(\beta z)}{\Theta(z)}$.

§III.5

2. (a) By (5.27) and (5.28), $T_n(f|[\alpha_N]_k) = n^{(k/2)-1} \sum_{a,b} f|[\alpha_N \sigma_a \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right)]_k$. Let $\alpha_{a,b} = \sigma_n \alpha_N \sigma_a \left(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix}\right) \alpha_N^{-1} \in \Delta^n(N, \{1\}, \mathbb{Z})$, and show that the $\alpha_{a,b}$ are in distinct cosets of $\Gamma_1(N)$ in $\Delta^n(N, \{1\}, \mathbb{Z})$, and hence form a set of representatives; so, by (5.27),

$(T_nf)|[\alpha_N]_k = n^{(k/2)-1}\,\Sigma_{a,b} f|[\alpha_{a,b}\alpha_N]_k = n^{(k/2)-1}\,\Sigma f|[\sigma_n\alpha_N\sigma_a(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix})]_k = T_n(f|[\alpha_N]_k)$, because $f|[\sigma_n]_k = f$. (b) $T_2F = 0$, $T_2\Theta^4 = \Theta^4 + 16F$, so $\{F, \frac{2}{3}F + \frac{1}{24}\Theta^4\}$ is a normalized eigenbasis. (c) $(T_2F)|[\alpha_4]_2 = 0 \neq T_2(F|[\alpha_4]_2)$. (d) Any eigenvector of T_2 or $[\alpha_4]_k$ must be an eigenvector of T_n; this includes F, $\frac{2}{3}F + \frac{1}{24}\Theta^4$, Θ^4, and so all of $M_2(\Gamma_0(4))$; by Proposition 40 applied to F, we have $\lambda_n = \sigma_1(n)$ (n odd). (e) Let $\Theta^4(z) = \Sigma\, a_nq^n$. For n odd, $a_n = \sigma_1(n)a_1 = 8\sigma_1(n)$ by Proposition 40. For $n = 2n_0$ twice an odd number, compare coefficients of q^{n_0} in $T_2\Theta^4 = \Theta^4 + 16F$ to get $a_n = a_{n_0} + 16\sigma_1(n_0) = 24\sigma_1(n_0)$. For $n = 2n_1$ divisible by 4, compare coefficients of q^{n_1} in $T_2\Theta^4 = \Theta^4 + 16F$ to get $a_n = a_{n_1} + 0 = 24\sigma_1(n_0)$ by induction. **3.** The first part follows from (5.19) with $n = 0$; a counterexample for other cusps is in Problem 2(b), since $\Theta^4 + 16F$ does not vanish at $s = -1/2$. **4.** (a) Both T_nf and $T_n'f$ are equal to $n^{(k/2)-1}\,\Sigma f|[\sigma_a(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix})]_k$ over the same a, b, d, because g.c.d.$(n, M) = 1$. Note that $g = M^{-k/2}f|[(\begin{smallmatrix} M & 0 \\ 0 & 1 \end{smallmatrix})]_k$. Thus, $M^{k/2}T_n'g = n^{(k/2)-1}\,\Sigma f|[(\begin{smallmatrix} M & 0 \\ 0 & 1 \end{smallmatrix})\sigma_a'(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix})]_k$, where $\sigma_a' \equiv (\begin{smallmatrix} 1/a & 0 \\ 0 & a \end{smallmatrix})$ mod MN. Set σ_a in the sum for T_nf equal to $(\begin{smallmatrix} M & 0 \\ 0 & 1 \end{smallmatrix})\sigma_a'(\begin{smallmatrix} M & 0 \\ 0 & 1 \end{smallmatrix})^{-1}$. Then $M^{k/2}T_n'g = (n^{(k/2)-1}\,\Sigma f|[\sigma_a(\begin{smallmatrix} M & 0 \\ 0 & 1 \end{smallmatrix})(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix})(\begin{smallmatrix} M & 0 \\ 0 & 1 \end{smallmatrix})^{-1}]_k)|[(\begin{smallmatrix} M & 0 \\ 0 & 1 \end{smallmatrix})]_k = (T_nf)|[(\begin{smallmatrix} M & 0 \\ 0 & 1 \end{smallmatrix})]_k$, because in $(\begin{smallmatrix} a & bM \\ 0 & d \end{smallmatrix}) = (\begin{smallmatrix} M & 0 \\ 0 & 1 \end{smallmatrix})(\begin{smallmatrix} a & b \\ 0 & d \end{smallmatrix})(\begin{smallmatrix} M & 0 \\ 0 & 1 \end{smallmatrix})^{-1}$ the entry bM runs through a complete set of residues mod d. (b) Since $S_8(\Gamma_0(2)) = \mathbb{C}f$, f is an eigenform for T_n; then by part (a), g is an eigenform with the same eigenvalue. (c) $T_2g = U_2V_2f = f$; $T_2f = -8f$ by a comparison of coefficients. Hence $\{f, 8g + f\}$ is the normalized eigenbasis.
5. $\langle T_pf, g\rangle = \Sigma_{j=0}^{p-1}\langle p^{(k/2)-1}f|[(\begin{smallmatrix} 1 & j \\ 0 & p \end{smallmatrix})]_k, g\rangle = \Sigma_{j=0}^{p-1}\langle f, p^{(k/2)-1}g|[(\begin{smallmatrix} p & -j \\ 0 & 1 \end{smallmatrix})]_k\rangle$ by (5.33). Since $g|[(\begin{smallmatrix} p & -j \\ 0 & 1 \end{smallmatrix})]_k = p^{k/2}g(pz - j) = p^{k/2}g(pz)$, we have $\langle T_pf, g\rangle = p\langle f, p^{k-1}g(pz)\rangle = p^k\langle f, V_pg\rangle$. **6.** (a) It has $a_1 = 0$.
(b) $f_1 = q^2 - 48q^3 + 8\cdot 27\cdot 5q^4 - 64\cdot 5\cdot 47q^5 + 4\cdot 9\cdot 5\cdot 17\cdot 47q^6 + \cdots$;
$f_2 = q - 24\cdot 43q^2 + 4\cdot 9\cdot 49\cdot 139q^3 + 64\cdot 171337q^4 + \cdots$;
$T_2f_1 = q + 1080q^2 + \cdots = f_2 + 2112f_1$;
$T_2f_2 = -24\cdot 43q + (64\cdot 171337 + 2^{23})q^2 + \cdots = -1032f_2 + 2^9 3^6 7^2 f_1$.
(c) Tr $T_2 = 1080$, Det $T_2 = -2^{10}\cdot 3^2\cdot 2221$; eigenvalues $= 540 \pm 12\sqrt{144169}$; normalized eigenforms are $f_2 + (131 \pm \sqrt{144169})12f_1$. (d) $2f_2 + 3144f_1$. (e) From part (d): $2(4\cdot 9\cdot 49\cdot 139) - 48\cdot 3144 = 8\cdot 9\cdot 5\cdot 23\cdot 41$. Alternately, compute $T_3f_1 = -48f_2 + 36\cdot 2619f_1$, $T_3f_2 = 36\cdot 6811f_2 + {*}f_1$ (we don't care about *), so Tr $T_3 = 36(2619 + 6811) = 8\cdot 9\cdot 5\cdot 23\cdot 41$. **8.** In case of difficulty, see §3.1(c) of [Serre 1973].

§IV.1

1. (a) $\rho\tilde{\gamma}\rho^{-1} = ((\begin{smallmatrix} m & 0 \\ 0 & 1 \end{smallmatrix}), m^{-1/4})((\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}), (\frac{c}{d})\varepsilon_d^{-1}\sqrt{cz + d})((\begin{smallmatrix} 1/m & 0 \\ 0 & 1 \end{smallmatrix}), m^{1/4}) = ((\begin{smallmatrix} a & mb \\ c/m & d \end{smallmatrix}), (\frac{c}{d})\varepsilon_d^{-1}\sqrt{cz/m + d})$ by (1.4). (b) $\gamma_1 = (\begin{smallmatrix} a & mb \\ c/m & d \end{smallmatrix})$, so $\tilde{\gamma}_1 = ((\begin{smallmatrix} a & mb \\ c/m & d \end{smallmatrix}), (\frac{c/m}{d})\varepsilon_d^{-1}\sqrt{cz/m + d}) = \rho\tilde{\gamma}\rho^{-1}((\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}), (\frac{m}{d}))$. Thus, they are equal if m is a square mod d. For m not a perfect square, to find γ for which $\tilde{\gamma}_1 = \rho\tilde{\gamma}\rho^{-1}(1, -1)$, it suffices to find d prime to $4m$ such that $(\frac{m}{d}) = -1$, since one can then find γ of the form $(\begin{smallmatrix} a & b \\ 4m & d \end{smallmatrix}) \in \Gamma_0(4m)$. To do this, first let $m = 2^\varepsilon m_1^2 m_0$, where $\varepsilon = 0$ or 1 and m_0 is squarefree. Choose $d \equiv 1$ mod 8 and $d \equiv d_0$ mod m_0, where d_0 is chosen so that $(\frac{d_0}{m_0}) = -1$. Then one easily checks that $(\frac{m}{d}) = (\frac{m_0}{d}) = (\frac{d}{m_0}) = -1$, as desired.
2. Replacing γ by $-\gamma$, if necessary, without loss of generality we may suppose that $b > 0$ or else $b = 0$ and $a > 0$. (a) $\rho\tilde{\gamma}\rho^{-1} = ((\begin{smallmatrix} 0 & -1 \\ N & 0 \end{smallmatrix}), N^{1/4}\sqrt{z})((\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}), (\frac{c}{d})\varepsilon_d^{-1}\sqrt{cz + d})\cdot((\begin{smallmatrix} 0 & 1/N \\ -1 & 0 \end{smallmatrix}), -iN^{1/4}\sqrt{z}) = ((\begin{smallmatrix} -c & -d \\ Na & Nb \end{smallmatrix}), N^{1/4}\sqrt{\frac{az+b}{cz+d}}(\frac{c}{d})\varepsilon_d^{-1}\sqrt{cz + d})((\begin{smallmatrix} 0 & 1/N \\ -1 & 0 \end{smallmatrix}), -iN^{1/4}\sqrt{z})$. Note that $\sqrt{(az + b)/(cz + d)}\cdot\sqrt{cz + d} = \eta_1\sqrt{az + b}$ with $\eta_1 = -1$ if $a < 0$ and $c \geq 0$, $\eta_1 = 1$ otherwise. Thus,

$\rho\tilde{\gamma}\rho^{-1} = \left(\left(\begin{smallmatrix} d & -c/N \\ -Nb & a \end{smallmatrix}\right), \eta_1 N^{1/4}\left(\frac{c}{d}\right)\varepsilon_d^{-1}\sqrt{a(-1/Nz)+b}\,(-iN^{1/4}\sqrt{z})\right) =$ $\left(\left(\begin{smallmatrix} d & -c/N \\ -Nb & a \end{smallmatrix}\right), \eta_1\left(\frac{c}{d}\right)\varepsilon_d^{-1}\sqrt{-Nbz+a}\right)$. (b) We have $\tilde{\gamma}_1 = \left(\left(\begin{smallmatrix} d & -c/N \\ -Nb & a \end{smallmatrix}\right), \left(\frac{-Nb}{a}\right)\varepsilon_a^{-1}\sqrt{-Nbz+a}\right)$. Since $ad - bc = 1$ and $4|c$, we have $a \equiv d \bmod 4$, and so $\varepsilon_a = \varepsilon_d$. Thus, it remains to compare $\left(\frac{-Nb}{a}\right)$ with $\eta_1\left(\frac{c}{d}\right)$. We suppose $c \geq 0$ (an analogous argument gives the same result if $c < 0$). Then $\left(\frac{Nb}{a}\right)\left(\frac{Nb}{d}\right) = \left(\frac{Nb}{1+bc}\right) = 1$, and $\left(\frac{-1}{a}\right) = \left(\frac{-1}{d}\right)$, so that $\left(\frac{-Nb}{a}\right) = \left(\frac{-Nb}{d}\right)$. Now $\left(\frac{-Nb}{d}\right)\left(\frac{c}{d}\right) = \left(\frac{N}{d}\right)\left(\frac{1-ad}{d}\right) = \left(\frac{N}{d}\right)\operatorname{sgn} a = \left(\frac{N}{d}\right)\eta_1$. Thus, $\tilde{\gamma}_1 = \rho\tilde{\gamma}\rho^{-1}\left(\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right), \left(\frac{N}{d}\right)\right)$. Again $\tilde{\gamma}_1$ and $\rho\tilde{\gamma}\rho^{-1}$ are equal if N is a perfect square; otherwise there exist γ for which the two differ by $(1, -1)$. **3.** At ∞, $\rho = 1$, $h = 1$, $t = 1$. At 0, take $\rho = \left(\left(\begin{smallmatrix} 0 & -1 \\ 1 & 0 \end{smallmatrix}\right), \sqrt{z}\right)$, so $\rho^{-1} = \left(\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right), -i\sqrt{z}\right)$, and we need $\tilde{\Gamma}_0(4) \ni \rho\left(\left(\begin{smallmatrix} 1 & h \\ 0 & 1 \end{smallmatrix}\right), t\right)\rho^{-1} =$ $\left(\left(\begin{smallmatrix} 0 & -1 \\ 1 & h \end{smallmatrix}\right), t\sqrt{z+h}\right)\left(\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right), -i\sqrt{z}\right) = \left(\left(\begin{smallmatrix} 1 & 0 \\ -h & 1 \end{smallmatrix}\right), -it\sqrt{-1/z+h}\cdot\sqrt{z}\right) =$ $\left(\left(\begin{smallmatrix} 1 & 0 \\ -h & 1 \end{smallmatrix}\right), -it\sqrt{hz-1}\right)$. This is in $\tilde{\Gamma}_0(4)$ if $h = 4$, $t = 1$. Finally, at the cusp $-\frac{1}{2}$ take $\alpha = \left(\begin{smallmatrix} 1 & 0 \\ -2 & 1 \end{smallmatrix}\right)$, $\rho = \left(\left(\begin{smallmatrix} 1 & 0 \\ -2 & 1 \end{smallmatrix}\right), \sqrt{-2z+1}\right)$, so $\rho^{-1} = \left(\left(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right), \sqrt{2z+1}\right)$. Since $\alpha\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right)\alpha^{-1} \in \Gamma_0(4)$, we can take $h = 1$. To find t, compute $\rho\left(\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right), t\right)\rho^{-1} =$ $\left(\left(\begin{smallmatrix} 3 & 1 \\ -4 & -1 \end{smallmatrix}\right), t\sqrt{-4z-1}\right)$, which is $j\left(\left(\begin{smallmatrix} 3 & 1 \\ -4 & -1 \end{smallmatrix}\right), z\right)$ provided that $t = i$.
5. (a) $S_{k/2}(\tilde{\Gamma}_0(4)) = 0$ if $k < 9$. For $k \geq 9$, it consists of elements of the form $\Theta F(16F - \Theta^4)P(\Theta, F)$, where P is a polynomial of pure weight $(k-9)/2$. Thus, for $k \geq 5$, $\dim S_{k/2}(\tilde{\Gamma}_0(4)) = [(k-5)/4]$. (b) Since $t = 1$ at the cusps ∞ and 0, there are always those two regular cusps. Since $t = i$ at the cusp $-\frac{1}{2}$, that cusp is k-regular if and only if $i^k = 1$, i.e., $4|k$. But $1 + [k/4] - [(k-5)/4] = 2$ if $k \geq 5$, $4\nmid k$ and $= 3$ if $k \geq 5$, $4|k$, as can be verified by checking for $k = 5, 6, 7, 8$ and then using induction to go from k to $k + 4$. (c) $\Theta F(\Theta^4 - 16F)(\Theta^4 - 2F)$. **6.** (a) If $d' \equiv d \bmod N$ and $N/4$ is odd, then $\left(\frac{N}{d}\right) = \left(\frac{N/4}{d}\right) = (-1)^{(N/4-1)(d-1)/2}\left(\frac{d}{N/4}\right) =$ $(-1)^{(N/4-1)(d'-1)/2}\left(\frac{d'}{N/4}\right) = \left(\frac{N}{d'}\right)$. If $N/4$ is even, then $d' \equiv d \bmod 8$, and the proof works the same way, with the additional observation that $\left(\frac{2}{d}\right) = \left(\frac{2}{d'}\right)$. (b) The proof that the cusp condition holds for $f|[\rho]_{k/2}$ is just like the analogous part of the proof of Proposition 17 in §III.3. Now let $\gamma = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right) \in \Gamma_0(N)$, and let $\gamma_1 = \left(\begin{smallmatrix} d & -c/N \\ -bN & a \end{smallmatrix}\right)$. By Problem 2 above, we have $\rho\tilde{\gamma}\rho^{-1} = (1, \chi_N(d))\tilde{\gamma}_1$. Thus, $(f|[\rho]_{k/2})|[\tilde{\gamma}]_{k/2} =$ $(f|[\rho\tilde{\gamma}\rho^{-1}]_{k/2})|[\rho]_{k/2} = \chi_N^k(d)(f|[\tilde{\gamma}_1]_{k/2})|[\rho]_{k/2} = \chi_N^k(d)\chi(a)f|[\rho]_{k/2}$. But since $ad \equiv 1 \bmod N$, we have $\chi(a) = \bar{\chi}(d)$. Thus, $(f|[\rho]_{k/2})|[\tilde{\gamma}]_{k/2} = \bar{\chi}\chi_N^k(d)f|[\rho]_{k/2}$, as desired. **7.** $\Theta(\infty) = 1$, $\Theta(-1/2) = 0$, $\Theta(0) = (1-i)/2$.

§IV.2

1. $E_{k/2}(\infty) = 1$, $E_{k/2}(0) = E_{k/2}(-1/2) = 0$; $F_{k/2}(0) = (i\sqrt{2})^{-k}$; $F_{k/2}(\infty) = F_{k/2}(-1/2) = 0$.
5. If one uses (2.16) and (2.19) rather than (2.17) and (2.20), then the solutions α and β to the resulting 2×2 equations involve χ, which depends on l. **6.** By Proposition 8, it suffices to show this for l squarefree. In that case use (2.16) and (2.19) to evaluate the l-th coefficient of $E_{k/2} + (1 + i^k)2^{-k/2}F_{k/2}$. **7.** $H_{k/2} =$ $\zeta(1-2\lambda) + \zeta(1-\lambda)q + \cdots$; $\zeta(1-\lambda) = 0$ if and only if $\lambda \geq 3$ is odd. **9.** Use Problem 1 above and Problem 7 of §IV.1 to find a and b so that $\Theta^k - aE_{k/2} - bF_{k/2}$ vanishes at the cusps. $a = 1$, $b = (1+i)^k 2^{-k/2}$. $\Theta^5 = E_{5/2} - (1+i)/\sqrt{2}F_{5/2}$, $\Theta^7 = E_{7/2} + (1-i)/\sqrt{2}F_{7/2}$.

§IV.3

2. The computation is almost identical to that in Problem 1 in §IV.1. **3.** (c) By the lemma in Proposition 43 in §III.5, right coset representatives for $\Gamma_1(N)$ modulo $\Gamma_1(N) \cap \alpha^{-1}\Gamma_1(N)\alpha$, where $\alpha = \left(\begin{smallmatrix} 1 & 0 \\ 0 & p^\nu \end{smallmatrix}\right)$, are $\alpha_b = \left(\begin{smallmatrix} 1 & b \\ 0 & p^\nu \end{smallmatrix}\right)$, $0 \leq b < p^\nu$. By Problems 1(b)

and 3(b), we use the α_b to get representatives for $\tilde{\Gamma}_1(N)$ modulo $\tilde{\Gamma}_1(N) \cap \xi_{p^\nu}^{-1} \tilde{\Gamma}_1(N) \xi_{p^\nu}$. Since $\alpha_b = \left(\begin{smallmatrix}1 & 0\\ 0 & p^\nu\end{smallmatrix}\right)\left(\begin{smallmatrix}1 & b\\ 0 & 1\end{smallmatrix}\right)$, we have $T_{p^\nu} f(z) = p^{\nu(k/4-1)} \cdot \Sigma_b f|[((\begin{smallmatrix}1 & 0\\ 0 & p^\nu\end{smallmatrix}), p^{\nu/4}) \cdot ((\begin{smallmatrix}1 & b\\ 0 & 1\end{smallmatrix}), 1)]_{k/2} = p^{-\nu} \Sigma_b f(\frac{z+b}{p^\nu}) = \Sigma_{p^\nu | n} a_n e^{2\pi i n z/p^\nu}$. The same argument works for T_{4^ν}, since the corresponding t is trivial even when $8 \nmid N$. **4.** (3.5) gives b_n for half integer weight $k/2$; in the case of integer weight k, the formula (5.19) of §III.5 with $m = p^2$ gives $b_n = a_{p^2 n} + \chi(p) p^{k-1} a_n + \chi(p^2) p^{2k-2} a_{n/p^2}$. If k is formally replaced by $k/2$ here, the middle term on the right differs by $\chi_{(-1)^\lambda n}(p)\sqrt{p}$ from the middle term on the right in (3.5). **5.** Let $\gamma = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) \in \Gamma_0(N)$ be such that $p^2 | b$. Then for $\gamma_j \in \Gamma_1(N)$ we have $\xi_{p^2} \tilde{\gamma}_j \tilde{\gamma} = \xi_{p^2} \tilde{\gamma} \xi_{p^2}^{-1} \xi_{p^2} \tilde{\gamma}^{-1} \tilde{\gamma}_j \tilde{\gamma}$. By Problem 2, we have $\xi_{p^2} \tilde{\gamma} \xi_{p^2}^{-1} = \tilde{\gamma}_1$ with $\gamma_1 = \left(\begin{smallmatrix}a & b/p^2\\ p^2 c & d\end{smallmatrix}\right) \in \Gamma_0(N)$. Let $\tau_j = \gamma^{-1} \gamma_j \gamma \in \Gamma_1(N)$. Then

$$\begin{aligned}(f|[\tilde{\Gamma}_1(N)\xi_{p^2}\tilde{\Gamma}_1(N)]_{k/2})|[\tilde{\gamma}]_{k/2} &= \sum_j f|[\xi_{p^2}\tilde{\gamma}_j\tilde{\gamma}]_{k/2} \\ &= \sum_j f|[\tilde{\gamma}_1 \xi_{p^2} \tilde{\tau}_j]_{k/2} = \chi(d) \sum_j f|[\xi_{p^2}\tilde{\tau}_j]_{k/2}.\end{aligned}$$

But one checks that the correspondence $\tilde{\Gamma}_1(N)\xi_{p^2}\tilde{\gamma}_j \mapsto \tilde{\Gamma}_1(N)\xi_{p^2}\tilde{\tau}_j$ permutes the right cosets in $\tilde{\Gamma}_1(N)\xi_{p^2}\tilde{\Gamma}_1(N)$, and so this equals $\chi(d) f|[\tilde{\Gamma}_1(N)\xi_{p^2}\tilde{\Gamma}_1(N)]_{k/2}$. (Verify that if τ_j and $\tau_{j'}$ are in the same right coset of $\Gamma_1(N) \cap \alpha^{-1}\Gamma_1(N)\alpha$ in $\Gamma_1(N)$, where $\alpha = \left(\begin{smallmatrix}1 & 0\\ 0 & p^2\end{smallmatrix}\right)$, then so are $\gamma_j = \gamma\tau_j\gamma^{-1}$ and $\gamma_{j'} = \gamma\tau_{j'}\gamma^{-1}$.) Thus, $T_{p^2} f|[\tilde{\gamma}]_{k/2} = \chi(d) T_{p^2} f$. It remains to note that $\Gamma_0(N)$ is generated by $\Gamma_1(N)$ and $\gamma = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) \in \Gamma_0(N)$ for which $p^2 | b$. Hence $T_{p^2} f|[\tilde{\gamma}]_{k/2} = \chi(d) T_{p^2} f$ for all $\gamma = \left(\begin{smallmatrix}a & b\\ c & d\end{smallmatrix}\right) \in \Gamma_0(N)$.

Bibliography

R. Alter, The congruent number problem, *Amer. Math. Monthly* **87** (1980), 43–45.

G. E. Andrews, *The Theory of Partitions*, Addison-Wesley, 1976.

N. Arthaud, On Birch and Swinnerton-Dyer's conjecture for elliptic curves with complex multiplication, *Compositio Math.* **37** (1978), 209–232.

E. Artin, *The Gamma Function*, Holt, Rinehart & Winston, 1964.

E. Artin, *Collected Papers*, Addison-Wesley, 1965.

A. G. van Asch, Modular forms of half integral weight, some explicit arithmetic, *Math. Annalen* **262** (1983), 77–89.

A. O. L. Atkin and J. Lehner, Hecke operators on $\Gamma_0(m)$, *Math. Annalen* **185** (1970), 134–160.

R. Bellman, *A Brief Introduction to Theta Functions*, Holt, Rinehart & Winston, 1961.

B. J. Birch, Conjectures on elliptic curves, *Amer. Math. Soc. Proc. Symp. Pure Math.* **8** (1963), 106–112.

B. J. Birch, Heegner points on elliptic curves, *Symp. Math., Ist. d. Alta Mat.* **15** (1975), 441–445.

B. J. Birch and H. P. F. Swinnerton-Dyer, Notes on elliptic curves I and II, *J. Reine Angew. Math.* **212** (1963), 7–25 and **218** (1965), 79–108.

Z. Borevich and I. Shafarevich, *Number Theory*, Academic Press, 1966.

A. Brumer and K. Kramer, The rank of elliptic curves, *Duke Math. J.* **44** (1977), 715–742.

J. Buhler, B. Gross, and D. Zagier, On the conjecture of Birch and Swinnerton-Dyer for ann elliptic curve of rank 3, *Math. Comp.* **44** (1985), 471–481.

D. Bump, S. Friedberg, and J. Hoffstein, A non-vanishing theorem for derivatives of automorphic L-functions with applications to elliptic curves, *Bull. Amer. Math. Soc.* **21** (1989), 89–93.

J. W. S. Cassels, Diophantine equations with special reference to elliptic curves, *J. London Math. Soc.* **41** (1966), 193–291.

J. W. S. Cassels and A. Fröhlich, eds., *Algebraic Number Theory*, Academic Press, 1967.

J. Coates, The arithmetic of elliptic curves with complex multiplication, *Proc. Int. Cong. Math. Helsinki* (1978), 351–355.

J. Coates, The work of Gross–Zagier on Heegner points and the derivatives of L-series, Sém. Bourbaki No. 635. In: *Astérisque* **133–134** (1986), 57–72.

J. Coates and A. Wiles, On the conjecture of Birch and Swinnerton-Dyer, *Inventiones Math.* **39** (1977), 223–251.

H. Cohen, Sommes de carrés, fontions L et formes modulaires, *C. R. Acad. Sc. Paris* **277** (1973), 827–830.

H. Cohen, Sums involving the values at negative integers of L-functions of quadratic characters, *Math. Annalen* **217** (1975), 271–285.

H. Cohen, Variations sur un thème de Siegel et Hecke, *Acta Arith.* **30** (1976), 63–93.

H. Cohen and J. Oesterlé, Dimensions des espaces de formes modulaires, Springer-Verlag Lecture Notes in Math. **627** (1976), 69–78.

H. Davenport and H. Hasse, Die Nullstellen der Kongruenz-zetafunktionen in gewissen zyklischen Fällen, *J. Reine Angew. Math.* **172** (1935), 151–182.

P. Deligne, Valeurs de fontions L et périodes d'intégrales, *Amer. Math. Soc. Proc. Symp. Pure Math.* **33** (1979), Part 2, 313–346.

P. Deligne and J.-P. Serre, Formes modulaires de poids 1, *Ann. Sci. Ecole Norm. Sup.* **7** (1974), 507–530.

L. E. Dickson, *History of the Theory of Numbers. Volume 2. Diophantine Analysis.* Chelsea, 1952.

P. G. L. Dirichlet, *Werke*, Chelsea, 1969.

B. Dwork, On the rationality of the zeta function, *Amer. J. Math.* **82** (1960), 631–648.

M. Eichler and D. Zagier, *The Theory of Jacobi Forms*, Birkhauser, 1984.

Y. Flicker, Automorphic forms on covering groups of $GL(2)$, *Inventiones Math.* **57** (1980), 119–182.

S. Gelbart, *Weil's Representation and the Spectrum of the Metaplectic Group*, Lecture Notes in Math. **530**, Springer-Verlag, 1976.

D. Goldfeld, Sur les produits eulériens attachés aux courbes elliptiques, *C. R. Acad. Sc. Paris* **294** (1982), 471–474.

D. Goldfeld, Gauss's class number problem for imaginary quadratic fields, *Bull. Amer. Math. Soc.* **13** (1985), 23–37.

D. Goldfeld, J. Hoffstein and S. J. Patterson, On automorphic functions of half-integral weight with applications to elliptic curves. In: *Number Theory Related to Fermat's Last Theorem*, Birkhauser, 1982, 153–193.

D. Goldfeld and C. Viola, Mean values of L-functions associated to elliptic, Fermat and other curves at the center of the critical strip, *J. Number Theory* **11** (1979), 305–320.

D. Goldfeld and C. Viola, Some conjectures on elliptic curves over cyclotomic fields, *Trans. Amer. Math. Soc.* **276** (1983), 511–515.

R. Greenberg, On the Birch and Swinnerton-Dyer conjectures, *Inventiones Math.* **72** (1983), 241–265.

B. H. Gross, *Arithmetic on Elliptic Curves with Complex Multiplication*, Lecture Notes in Math. **776**, Springer-Verlag, 1980.

B. H. Gross and D. Zagier, On the critical values of Hecke L-series, *Soc. Math. de France Mém.* No. 2 (1980), 49–54.

B. H. Gross and D. Zagier, Heegner points and derivatives of L-series, *Inventiones Math.* **84** (1986), 225–320.

B. H. Gross and D. Zagier, Points de Heegner et dérivées de fonctions L, *C. R. Acad. Sc. Paris* **297** (1983), 85–87.

R. C. Gunning, *Lectures on Modular Forms*, Princeton Univ. Press, 1962.

R. K. Guy, *Unsolved Problems in Number Theory*, Springer-Verlag, 1981.

G. H. Hardy, On the representation of a number as the sum of any number of squares, and in particular of five or seven, *Proc. Nat. Acad. Sci.* **4** (1918), 189–193. (*Collected Papers*, Vol. 1, 340, Clarendon Press, 1960.)

G. H. Hardy, On the representation of a number as the sum of any number of squares, and in particular of five, *Trans. Amer. Math. Soc.* **21** (1920), 255–284. (*Collected Papers*, Vol. I., 345, Clarendon Press, 1960.)

G. H. Hardy and E. M. Wright, *An Introduction to the Theory of Numbers*, 4th ed., Oxford Univ. Press, 1960.

R. Hartshorne, *Algebraic Geometry*, Springer-Verlag, 1977.

H. Hasse, *Number Theory*, Springer, 1980.
E. Hecke, Über die Bestimmung Dirichletscher Reihen durch ihre Funktionalgleichung, *Math. Annalen* **112** (1936), 664–699. (*Math. Werke*, 591–626.)
E. Hecke, Herleitung des Euler-Produktes der Zetafunktion und einiger L-Reihen aus ihrer Funktionalgleichung, *Math. Annalen* **119** (1944), 266–287. (*Math. Werke*, 919–940.)
E. Hecke, *Lectures on the Theory of Algebraic Numbers*, Springer-Verlag, 1981.
E. Hecke, *Lectures on Dirichlet Series, Modular Functions and Quadratic Forms*, Vandenhoeck and Ruprecht, 1983.
D. Husemöller, *Elliptic Curves*, Springer-Verlag, 1987.
K. Ireland and M. Rosen. *A Classical Introduction to Modern Number Theory*, 2nd ed. Springer-Verlag, 1990.
N. Katz, An overview of Deligne's proof of the Riemann hypothesis for varieties over finite fields, *Amer. Math. Soc. Proc. Symp. Pure Math.* **28** (1976), 275–305.
N. Katz, p-adic interpolation of real analytic Eisenstein series, *Annals of Math.* **104** (1976), 459–571.
N. Koblitz, Elliptic curve cryptosystems, *Math. Comp.* **48** (1987), 203–209.
N. Koblitz, *p-adic Numbers, p-adic Analysis, and Zeta-Functions*, 2nd ed., Springer-Verlag, 1984.
N. Koblitz, *p-adic Analysis: a Short Course on Recent Work*, Cambridge Univ. Press, 1980.
N. Koblitz, Why study equations over finite fields?, *Math. Magazine* **55** (1982), 144–149.
W. Kohnen, Modular forms of half integral weight on $\Gamma_0(4)$, *Math. Annalen* **248** (1980), 249–266.
W. Kohnen, Beziehungen zwischen Modulformen halbganzen Gewichts und Modulformen ganzen Gewichts, *Bonner Math. Scriften* **131** (1981).
W. Kohnen, Newforms of half-integral weight, *J. Reine und Angew. Math.* **333** (1982), 32–72.
W. Kohnen and D. Zagier, Values of L-series of modular forms at the center of the critical strip, *Inventiones Math.* **64** (1981), 175–198.
V. A. Kolyvagin, Finiteness of $E(\mathbf{Q})$ and Ш$(E, \mathbf{Q})$ for a class of Weil curves, *Math. of the USSR Izvestiya* **32** (1989), 523–542.
V. A. Kolyvagin, Euler systems, *The Grothendieck Festschrift* **II**, Birkhäuser, 1990, 435–483.
S. Lang, *Algebraic Number Theory*, Addison-Wesley, 1970.
S. Lang, *Elliptic Functions*, Addison-Wesley, 1973.
S. Lang. *Introduction to Modular Forms*, Springer-Verlag, 1976.
S. Lang, Sur la conjecture de Birch–Swinnerton-Dyer (d'après J. Coates et A. Wiles), Sém. Bourbaki No. 503. In: Springer-Verlag Lecture Notes in Math. **677** (1978), 189–200.
S. Lang, *Elliptic Curves Diophantine Analysis*, Springer-Verlag, 1978.
S. Lang, Units and class groups in number theory and algebraic geometry, *Bull Amer. Math. Soc.* **6** (1982), 253–316.
S. Lang, *Algebra*, Benjamin/Cummings, 1984.
W. J. LeVeque, *Fundamentals of Number Theory*, Addison-Wesley, 1977.
H. W. Lenstra, Jr., Factoring integers with elliptic curves, *Annals of Math.* **126** (1987), 649–673.
J. H. van Lint and G. van der Geer, *Introduction to Coding Theory and Algebraic Geometry*, Birkhäuser, 1988.
Yu. I. Manin, Cyclotomic fields and modular curves, *Russian Math. Surveys* **26** (1971), 7–78.
Yu. I. Manin, Modular forms and number theory, *Proc. Int. Cong. Math. Helsinki* (1978), 177–186.
D. Marcus, *Number Fields*, Springer-Verlag, 1977.

Modular Functions of One Variable IV, Lecture Notes in Math. **476**, Springer-Verlag, 1975.

A. Menezes and S. Vanstone, The implementation of elliptic curve cryptosystems, *Advances in Cryptology—Auscrypt '90*, Springer-Verlag, 1990, 2–13.

V. Miller, Use of elliptic curves in cryptography, *Advances in Cryptology—Crypto '85*, Springer-Verlag, 1986, 417–426.

C.J. Moreno, The higher reciprocity laws: an example, *J. Number Theory* **12** (1980), 57–70.

D. Mumford, *Algebraic Geometry I: Complex Projective Varieties*, Springer-Verlag, 1976.

M.R. Murty and V. K. Murty, Mean values of derivatives of modular *L*-series, *Annals of Math.* **133** (1991), 447–475.

S. Niwa, Modular forms of half-integral weight and the integral of certain theta-functions, *Nagoya Math. J.* **56** (1975), 147–161.

J. Oesterlé, Nouvelles approaches du "Théorème" de Fermat, Sém. Bourbaki No. 694. In: *Astérisque* **161–162** (1988), 165–186.

A. Ogg, *Modular Forms and Dirichlet Series*, W. A. Benjamin, 1969.

R. A. Rankin, *Modular Forms and Functions*, Cambridge Univ. Press, 1977.

K. Rubin, Tate–Shafarevich groups and *L*-functions of elliptic curves with complex multiplication, *Inventiones Math.* **89** (1987), 527–560.

K. Rubin, The work of Kolyvagin on the arithmetic of elliptic curves, Spring-Verlag Lecture Notes in Math. **1399** (1989), 128–136.

B. Schoeneberg, *Elliptic Modular Functions: an Introduction*, Springer-Verlag, 1974.

J.-P. Serre, Formes modulaires et fonctions zêta *p*-adiques, Springer-Verlag Lecture Notes in Math. **350** (1973), 191–268.

J.-P. Serre, *A Course in Arithmetic*, Springer-Verlag, 1977.

J.-P. Serre and H. M. Stark, Modular forms of weight 1/2, Springer-Verlag Lecture Notes in Math. **627** (1977), 27–67.

I. R. Shafarevich, *Basic Algebraic Geometry*, Springer-Verlag, 1977.

G. Shimura, *Introduction to the Arithmetic Theory of Automorphic Functions*, Princeton Univ. Press, 1971.

G. Shimura, On modular forms of half-integral weight, *Annals of Math.* **97** (1973), 440–481.

G. Shimura, Modular forms of half integral weight, Springer-Verlag Lecture Notes in Math. **320** (1973), 59–74.

T. Shintani, On construction of holomorphic cusp forms of half-integral weight, *Nagoya Math. J.* **58** (1975), 83–126.

C. L. Siegel, *On Advanced Analytic Number Theory*, Tata Institute (Bombay), 1961.

J. Silverman, *The Arithmetic of Elliptic Curves*, Springer-Verlag, 1986.

N. M. Stephens, Congruence properties of congruent numbers, *Bull. London Math. Soc.* **7** (1975), 182–184.

H. P. F. Swinnerton-Dyer, The conjectures of Birch and Swinnerton-Dyer and of Tate. In: *Proc. Conf. Local Fields*, Springer-Verlag, 1967.

J. Tate, The arithmetic of elliptic curves, *Inventiones Math.* **23** (1974), 179–206.

J. Tunnell, A classical Diophantine problem and modular forms of weight 3/2, *Inventiones Math.* **72** (1983), 323–334.

M.-F. Vignéras, Valeur au centre de symmétrie des fonctions L associées aux formes modulaires, Sém. Delange–Pisot–Poitou, 1980.

B. L. van der Waerden, *Algebra* (2 vols.), Frederick Ungar, 1970.

J. L. Waldspurger, Correspondance de Shimura, *J. Math. Pures et Appl.* **59** (1980), 1–132.

J. L. Waldspurger, Sur les coefficients de Fourier des formes modulaires de poids demi-entier, *J. Math. Pures et Appl.* **60** (1981), 375–484.

R. J. Walker, *Algebraic Curves*, Springer-Verlag, 1978.

L. Washington, *Introduction to Cyclotomic Fields*, Springer-Verlag, 1982.

A. Weil, Number of solutions of equations in finite fields, *Bull. Amer. Math. Soc.* **55** (1949), 497–508. (*Collected Papers*, Vol. I, 399–410.)

A. Weil, Jacobi sums as "Grössencharaktere," *Trans. Amer. Math. Soc.* **73** (1952), 487–495. (*Collected Papers*, Vol. II, 63–71.)

A. Weil, On a certain type of characters of the idèle-class group of an algebraic number-field, *Proc. Intern. Symp. on Alg. Num. Theory*, Tokyo-Nikko (1955), 1–7. (*Collected Papers*, Vol. II, 255–261.)

A. Weil, Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen, *Math. Annalen* **168** (1967), 149–156.

A. Weil, Review of "The mathematical career of Pierre de Fermat" by M. S. Mahoney, *Bull Amer. Math. Soc.* **79** (1973), 1138–1149. (*Collected Papers*, Vol. III, 266–277.)

A. Weil, *Basic Number Theory*, 3rd ed., Springer-Verlag, 1974.

A. Weil, Sur les sommes de trois et quatre carrés, *L'Enseignement Math.* **20** (1974), 215–222. (*Collected Papers*, Vol. III, 303–310.)

A. Weil, La cyclotomie jadis et naguère, Sém. Bourbaki No. 452. In: Springer-Verlag Lecture Notes in Math. **431** (1975), 318–338 and in *l'Enseignement Math.* **20** (1974), 247–263. (*Collected Papers*, Vol. III, 311–327.)

A. Weil, Sommes de Jacobi et caractères de Hecke, *Gött. Nachr. Nr. 1* (1974), 14 pp. (*Collected Papers*, Vol. III, 329–342.)

A. Weil, *Number Theory: an Approach Through History from Hammurapi to Legendre*, Birkhauser, 1983.

E. Whittaker and G. Watson, *A Course of Modern Analysis*, 4th ed., Cambridge Univ. Press, 1958.

D. Zagier, *L*-series of elliptic curves, the Birch–Swinnerton-Dyer conjecture, and the class number problem of Gauss, *Notices Amer. Math. Soc.* **31** (1984), 739–743.

D. Zagier, Nombres de classes et formes modulaires de poids 3/2, *C. R. Acad. Sc. Paris* **281** (1975), 883–886.

D. Zagier, Modular forms associated to real quadratic fields, *Inventiones Math.* **30** (1975), 1–46.

D. Zagier, On the values at negative integers of the zeta function of real quadratic fields, *Enseignement Math.* **22** (1976), 55–95.

D. Zagier, Modular forms whose Fourier coefficients involve zeta-functions of quadratic fields, Springer Lecture Notes in Math. **627** (1977), 105–169.

Index

Graduate Texts in Mathematics

1 TAKEUTI/ZARING. Introduction to Axiomatic Set Theory. 2nd ed.
2 OXTOBY. Measure and Category. 2nd ed.
3 SCHAEFER. Topological Vector Spaces. 2nd ed.
4 HILTON/STAMMBACH. A Course in Homological Algebra. 2nd ed.
5 MAC LANE. Categories for the Working Mathematician. 2nd ed.
6 HUGHES/PIPER. Projective Planes.
7 SERRE. A Course in Arithmetic.
8 TAKEUTI/ZARING. Axiomatic Set Theory.
9 HUMPHREYS. Introduction to Lie Algebras and Representation Theory.
10 COHEN. A Course in Simple Homotopy Theory.
11 CONWAY. Functions of One Complex Variable I. 2nd ed.
12 BEALS. Advanced Mathematical Analysis.
13 ANDERSON/FULLER. Rings and Categories of Modules. 2nd ed.
14 GOLUBITSKY/GUILLEMIN. Stable Mappings and Their Singularities.
15 BERBERIAN. Lectures in Functional Analysis and Operator Theory.
16 WINTER. The Structure of Fields.
17 ROSENBLATT. Random Processes. 2nd ed.
18 HALMOS. Measure Theory.
19 HALMOS. A Hilbert Space Problem Book. 2nd ed.
20 HUSEMOLLER. Fibre Bundles. 3rd ed.
21 HUMPHREYS. Linear Algebraic Groups.
22 BARNES/MACK. An Algebraic Introduction to Mathematical Logic.
23 GREUB. Linear Algebra. 4th ed.
24 HOLMES. Geometric Functional Analysis and Its Applications.
25 HEWITT/STROMBERG. Real and Abstract Analysis.
26 MANES. Algebraic Theories.
27 KELLEY. General Topology.
28 ZARISKI/SAMUEL. Commutative Algebra. Vol.I.
29 ZARISKI/SAMUEL. Commutative Algebra. Vol.II.
30 JACOBSON. Lectures in Abstract Algebra I. Basic Concepts.
31 JACOBSON. Lectures in Abstract Algebra II. Linear Algebra.
32 JACOBSON. Lectures in Abstract Algebra III. Theory of Fields and Galois Theory.
33 HIRSCH. Differential Topology.
34 SPITZER. Principles of Random Walk. 2nd ed.
35 ALEXANDER/WERMER. Several Complex Variables and Banach Algebras. 3rd ed.
36 KELLEY/NAMIOKA et al. Linear Topological Spaces.
37 MONK. Mathematical Logic.
38 GRAUERT/FRITZSCHE. Several Complex Variables.
39 ARVESON. An Invitation to C^*-Algebras.
40 KEMENY/SNELL/KNAPP. Denumerable Markov Chains. 2nd ed.
41 APOSTOL. Modular Functions and Dirichlet Series in Number Theory. 2nd ed.
42 SERRE. Linear Representations of Finite Groups.
43 GILLMAN/JERISON. Rings of Continuous Functions.
44 KENDIG. Elementary Algebraic Geometry.
45 LOÈVE. Probability Theory I. 4th ed.
46 LOÈVE. Probability Theory II. 4th ed.
47 MOISE. Geometric Topology in Dimensions 2 and 3.
48 SACHS/WU. General Relativity for Mathematicians.
49 GRUENBERG/WEIR. Linear Geometry. 2nd ed.
50 EDWARDS. Fermat's Last Theorem.
51 KLINGENBERG. A Course in Differential Geometry.
52 HARTSHORNE. Algebraic Geometry.
53 MANIN. A Course in Mathematical Logic.
54 GRAVER/WATKINS. Combinatorics with Emphasis on the Theory of Graphs.
55 BROWN/PEARCY. Introduction to Operator Theory I: Elements of Functional Analysis.
56 MASSEY. Algebraic Topology: An Introduction.
57 CROWELL/FOX. Introduction to Knot Theory.
58 KOBLITZ. p-adic Numbers, p-adic Analysis, and Zeta-Functions. 2nd ed.
59 LANG. Cyclotomic Fields.
60 ARNOLD. Mathematical Methods in Classical Mechanics. 2nd ed.
61 WHITEHEAD. Elements of Homotopy Theory.
62 KARGAPOLOV/MERLZJAKOV. Fundamentals of the Theory of Groups.
63 BOLLOBAS. Graph Theory.
64 EDWARDS. Fourier Series. Vol. I. 2nd ed.
65 WELLS. Differential Analysis on Complex Manifolds. 2nd ed.

Graduate Texts in Mathematics

66 WATERHOUSE. Introduction to Affine Group Schemes.
67 SERRE. Local Fields.
68 WEIDMANN. Linear Operators in Hilbert Spaces.
69 LANG. Cyclotomic Fields II.
70 MASSEY. Singular Homology Theory.
71 FARKAS/KRA. Riemann Surfaces. 2nd ed.
72 STILLWELL. Classical Topology and Combinatorial Group Theory. 2nd ed.
73 HUNGERFORD. Algebra.
74 DAVENPORT. Multiplicative Number Theory. 3rd ed.
75 HOCHSCHILD. Basic Theory of Algebraic Groups and Lie Algebras.
76 IITAKA. Algebraic Geometry.
77 HECKE. Lectures on the Theory of Algebraic Numbers.
78 BURRIS/SANKAPPANAVAR. A Course in Universal Algebra.
79 WALTERS. An Introduction to Ergodic Theory.
80 ROBINSON. A Course in the Theory of Groups. 2nd ed.
81 FORSTER. Lectures on Riemann Surfaces.
82 BOTT/TU. Differential Forms in Algebraic Topology.
83 WASHINGTON. Introduction to Cyclotomic Fields. 2nd ed.
84 IRELAND/ROSEN. A Classical Introduction to Modern Number Theory. 2nd ed.
85 EDWARDS. Fourier Series. Vol. II. 2nd ed.
86 VAN LINT. Introduction to Coding Theory. 2nd ed.
87 BROWN. Cohomology of Groups.
88 PIERCE. Associative Algebras.
89 LANG. Introduction to Algebraic and Abelian Functions. 2nd ed.
90 BRØNDSTED. An Introduction to Convex Polytopes.
91 BEARDON. On the Geometry of Discrete Groups.
92 DIESTEL. Sequences and Series in Banach Spaces.
93 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part I. 2nd ed.
94 WARNER. Foundations of Differentiable Manifolds and Lie Groups.
95 SHIRYAEV. Probability. 2nd ed.
96 CONWAY. A Course in Functional Analysis. 2nd ed.
97 KOBLITZ. Introduction to Elliptic Curves and Modular Forms. 2nd ed.
98 BRÖCKER/TOM DIECK. Representations of Compact Lie Groups.
99 GROVE/BENSON. Finite Reflection Groups. 2nd ed.
100 BERG/CHRISTENSEN/RESSEL. Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions.
101 EDWARDS. Galois Theory.
102 VARADARAJAN. Lie Groups, Lie Algebras and Their Representations.
103 LANG. Complex Analysis. 3rd ed.
104 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part II.
105 LANG. $SL_2(\mathbf{R})$.
106 SILVERMAN. The Arithmetic of Elliptic Curves.
107 OLVER. Applications of Lie Groups to Differential Equations. 2nd ed.
108 RANGE. Holomorphic Functions and Integral Representations in Several Complex Variables.
109 LEHTO. Univalent Functions and Teichmüller Spaces.
110 LANG. Algebraic Number Theory.
111 HUSEMÖLLER. Elliptic Curves.
112 LANG. Elliptic Functions.
113 KARATZAS/SHREVE. Brownian Motion and Stochastic Calculus. 2nd ed.
114 KOBLITZ. A Course in Number Theory and Cryptography. 2nd ed.
115 BERGER/GOSTIAUX. Differential Geometry: Manifolds, Curves, and Surfaces.
116 KELLEY/SRINIVASAN. Measure and Integral. Vol. I.
117 SERRE. Algebraic Groups and Class Fields.
118 PEDERSEN. Analysis Now.
119 ROTMAN. An Introduction to Algebraic Topology.
120 ZIEMER. Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation.
121 LANG. Cyclotomic Fields I and II. Combined 2nd ed.
122 REMMERT. Theory of Complex Functions. *Readings in Mathematics*
123 EBBINGHAUS/HERMES et al. Numbers. *Readings in Mathematics*
124 DUBROVIN/FOMENKO/NOVIKOV. Modern Geometry—Methods and Applications. Part III.
125 BERENSTEIN/GAY. Complex Variables: An Introduction.
126 BOREL. Linear Algebraic Groups. 2nd ed.
127 MASSEY. A Basic Course in Algebraic Topology.
128 RAUCH. Partial Differential Equations.
129 FULTON/HARRIS. Representation Theory: A First Course. *Readings in Mathematics*
130 DODSON/POSTON. Tensor Geometry.

131 LAM. A First Course in Noncommutative Rings.
132 BEARDON. Iteration of Rational Functions.
133 HARRIS. Algebraic Geometry: A First Course.
134 ROMAN. Coding and Information Theory.
135 ROMAN. Advanced Linear Algebra.
136 ADKINS/WEINTRAUB. Algebra: An Approach via Module Theory.
137 AXLER/BOURDON/RAMEY. Harmonic Function Theory. 2nd ed.
138 COHEN. A Course in Computational Algebraic Number Theory.
139 BREDON. Topology and Geometry.
140 AUBIN. Optima and Equilibria. An Introduction to Nonlinear Analysis.
141 BECKER/WEISPFENNING/KREDEL. Gröbner Bases. A Computational Approach to Commutative Algebra.
142 LANG. Real and Functional Analysis. 3rd ed.
143 DOOB. Measure Theory.
144 DENNIS/FARB. Noncommutative Algebra.
145 VICK. Homology Theory. An Introduction to Algebraic Topology. 2nd ed.
146 BRIDGES. Computability: A Mathematical Sketchbook.
147 ROSENBERG. Algebraic K-Theory and Its Applications.
148 ROTMAN. An Introduction to the Theory of Groups. 4th ed.
149 RATCLIFFE. Foundations of Hyperbolic Manifolds.
150 EISENBUD. Commutative Algebra with a View Toward Algebraic Geometry.
151 SILVERMAN. Advanced Topics in the Arithmetic of Elliptic Curves.
152 ZIEGLER. Lectures on Polytopes.
153 FULTON. Algebraic Topology: A First Course.
154 BROWN/PEARCY. An Introduction to Analysis.
155 KASSEL. Quantum Groups.
156 KECHRIS. Classical Descriptive Set Theory.
157 MALLIAVIN. Integration and Probability.
158 ROMAN. Field Theory.
159 CONWAY. Functions of One Complex Variable II.
160 LANG. Differential and Riemannian Manifolds.
161 BORWEIN/ERDÉLYI. Polynomials and Polynomial Inequalities.
162 ALPERIN/BELL. Groups and Representations.
163 DIXON/MORTIMER. Permutation Groups.
164 NATHANSON. Additive Number Theory: The Classical Bases.
165 NATHANSON. Additive Number Theory: Inverse Problems and the Geometry of Sumsets.
166 SHARPE. Differential Geometry: Cartan's Generalization of Klein's Erlangen Program.
167 MORANDI. Field and Galois Theory.
168 EWALD. Combinatorial Convexity and Algebraic Geometry.
169 BHATIA. Matrix Analysis.
170 BREDON. Sheaf Theory. 2nd ed.
171 PETERSEN. Riemannian Geometry.
172 REMMERT. Classical Topics in Complex Function Theory.
173 DIESTEL. Graph Theory. 2nd ed.
174 BRIDGES. Foundations of Real and Abstract Analysis.
175 LICKORISH. An Introduction to Knot Theory.
176 LEE. Riemannian Manifolds.
177 NEWMAN. Analytic Number Theory.
178 CLARKE/LEDYAEV/STERN/WOLENSKI. Nonsmooth Analysis and Control Theory.
179 DOUGLAS. Banach Algebra Techniques in Operator Theory. 2nd ed.
180 SRIVASTAVA. A Course on Borel Sets.
181 KRESS. Numerical Analysis.
182 WALTER. Ordinary Differential Equations.
183 MEGGINSON. An Introduction to Banach Space Theory.
184 BOLLOBAS. Modern Graph Theory.
185 COX/LITTLE/O'SHEA. Using Algebraic Geometry.
186 RAMAKRISHNAN/VALENZA. Fourier Analysis on Number Fields.
187 HARRIS/MORRISON. Moduli of Curves.
188 GOLDBLATT. Lectures on the Hyperreals: An Introduction to Nonstandard Analysis.
189 LAM. Lectures on Modules and Rings.
190 ESMONDE/MURTY. Problems in Algebraic Number Theory.
191 LANG. Fundamentals of Differential Geometry.
192 HIRSCH/LACOMBE. Elements of Functional Analysis.
193 COHEN. Advanced Topics in Computational Number Theory.
194 ENGEL/NAGEL. One-Parameter Semigroups for Linear Evolution Equations.
195 NATHANSON. Elementary Methods in Number Theory.
196 OSBORNE. Basic Homological Algebra.
197 EISENBUD/HARRIS. The Geometry of Schemes.
198 ROBERT. A Course in p-adic Analysis.
199 HEDENMALM/KORENBLUM/ZHU. Theory of Bergman Spaces.
200 BAO/CHERN/SHEN. An Introduction to Riemann–Finsler Geometry.

201 HINDRY/SILVERMAN. Diophantine Geometry: An Introduction.
202 LEE. Introduction to Topological Manifolds.
203 SAGAN. The Symmetric Group: Representations, Combinatorial Algorithms, and Symmetric Functions. 2nd ed.
204 ESCOFIER. Galois Theory.
205 FÉLIX/HALPERIN/THOMAS. Rational Homotopy Theory.
206 MURTY. Problems in Analytic Number Theory.
Readings in Mathematics
207 GODSIL/ROYLE. Algebraic Graph Theory.